Paired ordered responses -- counted
(true contingency tables)

		Exh.	Page
Introduction 29x-1			A
-- contingency table, at 29x-1		1	29x-2/3
			B
29xA.	Two-by-two tables 29x-1	2	29x-9/10
	-- ff-log or double flog, at 29x-4	3	29x-11
	- double-flogs by hand 29x-4	4	29x-13/16
	- the standard breakdown 29x-5	5	29x-17/18
			C
29xB.	Two ordered responses -- school girls 29x-7	6	29x-22/23
	- some are weaker 29x-12	7	29x-24/25
		8	29x-27
29xC.	Extending the analysis 29x-20	9	29x-29/30
	- a flog product carrier 29x-20		D
	-- centering, at 29x-20	10	29x-32/33
	- looking harder at our fit 29x-26	11	29x-35
		12	29x-36/38
29xD.	The smallest corners again -- 3 examples 29x-31	13	29x-39
	- men's vision 29x-31	14	29x-40/42
	- call money 29x-34	15	29x-44/46
	- bean dimensions 29x-34		E
	- discussion 29x-43	16	29x-49
		17	29x-50
29xE.	Theoretical distributions 29x-47	18	29x-51
	- double flogs for biGaussians 29x-54	19	29x-52
	- residuals after log smallest corner 29x-54	20	29x-53
	- deeper residuals 29x-54	21	29x-55
		22	29x-56
29xF.	The bean dimensions further fitted 29x-67	23	29x-57
	- the 2nd step 29x-68	24	29x-58
	- the 3rd step 29x-71	25	29x-59
	- the 4th step 29x-75	26	29x-60
	- the 5th step 29x-75	27	29x-61/63
	- the 6th step 29x-82	28	29x-64
	- strays? 29x-82	29	29x-66
	- comparison of steps 29x-87		F
		30	29x-69/70
29xG.	Further analysis of other examples 29x-89	31	29x-72/73
	- call money 29x-89	32	29x-74
	- school-girls 29x-95	33	29x-76
		34	29x-77
	**********	35	29x-78/79
		36	29x-80/81
29xH.	How far have we come? 29x-97	37	29x-83/84
		38	29x-85
	**********	39	29x-86
		40	29x-88
29xP.	Problems for chapter 29x 29x-102		G
		41	29x-90
	**********	42	29x-91/94
		43	29x-96
29xR.	Technical references 29x-104		H
			P
			R

29x

Paired ordered responses -- counted (true contingency tables)

As we have begun to see, there are many kinds of two-way tables of counts. We are only now about to look at the kind of two-way table of counts that has sometimes been regarded as the only kind. When someone writes or speaks of a

contingency table

the reasonable supposition is that he is referring to a two-way (or more-way) table of counts in which

both factors conceal responses.

The word "contingency" here refers to the "happening together" of the two responses.

As two-way tables in which both factors are of the same character, we ought to expect a break-down of contingency tables into parts of different kinds, each kind of part having a different kind of interpretation. This we will learn to do easily when both factors are ordered -- when their versions come in a natural order.

30xA. <u>Two-by-two tables</u>

In 1906, Karl Pearson published a study of the relationship of intelligence to various physical characteristics in school children, from which we draw exhibit 1. Forming the difference of two f-logs in each of the two ways of course gives the same

exhibit 1 of chapter 29x -- school boys

Intelligence and athletic power in school boys

A) The DATA -- counts

	Intelligent	Not Intelligent	(sum)
Athletic	581	567	(1148)
Not Athletic	209	351	(560)
(sum)	(790)	(918)	

B) One DIFFERENCE of f-LOGS -- in matched form, see exhibit 1 of chapter 28

	Intelligent	Not Intelligent
% Athletic	73.5%	61.8%
f-log raw	.44	.21

Difference in f-logs = .23

C) The other DIFFERENCE of f-LOGS -- in matched form

	% Intelligent	f-log (raw)
Athletic	50.7%	.01
Non-athletic	37.4%	-.22

Difference in f-logs = .23

D) EITHER DIFFERENCE of f-LOGS -- still in matched form

$$= 1.15 \log 581 - 1.15 \log 567 - 1.15 \log 209 + 1.15 \log 351$$

$$= 1.15 \log \frac{581 \times 351}{567 \times 209}$$

We will call this a

ff-log or double f-log

to reflect folding both ways. (This is a matched ff-log. It would be a raw ff-log if we omitted all factors of 1.15.)

exhibit 1 (cont'd.)

E) ALTERNATIVE CALCULATION by HAND

matched

$$\left[\text{ff-log} = 1.15 \log \frac{581\times351}{567\times209}\right]$$

$$\text{ff-log} = 1.15 \log \frac{f^*}{1 - f^*}$$

raw

$$\text{ff-log} = \log \frac{581\times351}{567\times209}$$

or

$$\left[\text{ff-log} = \log \frac{f^*}{1-f^*}\right]$$

where

$$f^* = \frac{581\times351}{567\times209 + 581\times351}$$

the matched double f-log is thus calculable by one reference to exhibit 1 of chapter 28.

F) DATA SOURCE

K. Pearson, 1906. On the relationship of intelligence to size and shape of head, and to other physical and mental characters. Biometrika 5:105-146. Condensed from table xxxiii on page 144.

answer, since each can be put in the symmetric form of panel D. As that panel states, we are defining this symmetric quantity as the

ff-log or double flog

(either read and pronounced "double-flog") for any 2-by-2 table of counts. Here the double f's reflect the double foldings -- one horizontal and one vertical -- which we can think of as involved in its definition.

When we write or say "double-flog" for ff-log we are likely to write or say "single-flog" for f-log. We abbreviate away the hyphen freely, using "flog" and "fflog" in exhibits and formulas.

We can calculate a double flog as the difference of whichever pair of single flogs it is convenient to calculate. We may also, if we wish, calculate it as in panel E of exhibit 1.

* double-flogs by hand *

Since calculating logs without a table is rarely undertaken, a man who is calculating double flogs will ordinarily use either some sort of table or a slide rule.

At the cost of losing a small amount of precision, the quickest use of a table is to use exhibit 3 of chapter 3 to find a 2-decimal log for each of the four counts and then add and subtract together these four logs. The matching factor is then forgotten. For the example of exhibit 1, this gives

2.76 - 2.75 - 2.32 + 2.55 = .24, differing by .01 from the more precise value of .23.

A little more precision can be had by the route of panel E of exhibit 1, where two multiplications and one division are needed as a preparation for the use of exhibit 1 of chapter 28. The matching factor is now automatically included, so we usually leave it in.

More precision and less effort is available by use of a slide rule to (a) do the multiplications and divisions -- most conveniently, perhaps, as one count divided by another, then multiplied by the third, then divided by the fourth -- and (b) in a separate operation, to look up the log of the result. Here the 1.15 matching factor is most easily omitted.

For still more precision -- rarely needed -- or when a slide rule is not available, we can use a log table -- preferably a small one, getting 3 (or perhaps 4) decimals in our logs. Again looking up all four logs is the simplest approach. Again the matching factor is to be omitted.

* the standard breakdown *

Let us turn back to exhibit 1, and ask what sort of a breakdown we are making for this table. Our breakdown is of the nicest sort -- we start with four counts and come out with an equal number of answers.

For the example of exhibit 1, the four answers are:

- the total number classified, here 1708.

- the fraction intelligent, here 790/1708 = 46.2% -- probably well-enough described by a flog of -.08 (in matched form).

- the fraction athletic, here 1148/1708 = 67.2% -- probably well-enough described by a flog of +.36 (in matched form).

- the fflog of .27 (in matched form).

These answers are of three kinds -- we may be equally interested in all, but we are more likely to be much more interested in some than in others.

Our interest in the total is likely to be quite limited. It does tell us about how much data has been collected, something that might be of greater interest were we doing confirmatory data analysis. Occasionally this part is of real interest by itself.

The two fractions, probably given as flogs, may or may not be of interest for themselves. If one classification were sex, for example, whose two versions are well known to all, then the corresponding fraction might well be interesting for itself. In the present example, the fractions are of little interest themselves, at least without a lot of explanation. Just where was the line drawn between intelligent and not

intelligent? Between athletic and not athletic? Without answers to such questions, the fractions have little meaning in themselves.

These fractions always help, of course, to give a background against which to think of the double-flog. We will, in particular, feel rather differently about a double-flog of a certain size if one or both fractions are extreme.

Finally, and most importantly for us here, there is the double flog itself. By its difference from zero it tells us about the direction and amount of the failure of the four counts to be proportional -- if the failure of the two responses to be "independent" in the sense of section 30B, so far as our 2-by-2 table shows. It is the only one of the four parts which requires having all four counts and making a 2-by-2 table analysis. Thus it will often have the lion's share of our attention when we make two-way analyses.

29xB. <u>Two ordered responses -- school girls</u>

Tables occur rather frequently in which both factors are ordered responses (see chapter 29 for background) and deserve equal treatment. The simplest case of course consists of 2-by-2 tables, just treated. (For a 2-by-2 table the question of order can always be bypassed.)

Larger tables of this sort can be treated along the same lines -- again using double-flogs. We have only to condense them to 2-by-2 tables, using all possible pairs of cutting

points, and then to calculate all the corresponding double flogs. We can begin with the uncondensed form of the data which provided exhibit 1. This appears, with its analysis, in exhibit 2. As panel E -- of exhibit 2 -- shows, we have two reasonable alternative descriptions of the ff-logs:

- near .31 for high cuts, but near .48 for low ones. (Less plausible; better fit.)
- near a straight line rising from .25 for the highest cut to .50 for the lowest. (Poorer fit; more plausible.)

In either case, strength of relationship seems to increase with intelligence.

We leave the reader to ponder over his or her choice.

The essential point is that we have put the interesting part of the data in a form where we can take a hard look at its behavior. If double-flogs are a good way to express what is going on -- something we do not yet know for this example -- then we see, in particular, an increased relation between lack of intellectual power and lack of athletic power. (Without looking hard at much more extensive data, we would not know whether this is "just engineering". There may be a better way -- a more scientific way -- to look at the data -- one in which the degree of apparent relationship is more constant.)

exhibit 2 of chapter 29x -- school boys

Intellectual power and athletic power in school boys
(uncondensed)

A) The DATA

Teacher's rating	Cut	Athletic	Non-athletic	(sum)
Quick intelligent		159.5	46	205.5
	α			
Intelligent		421.75	163.25	585
	β			
Slow intelligent		355.5	187.5	543
	γ			
Slow		158.75	99.75	258.5
	δ			
Slow dull		40.5	48.5	89
	ε			
Very dull		12	15	27
(sum)		1148	560	

B) CUMULATIONS ABOVE CUTS

Cut	Athletic	Non-athletic	(Sum)	f-log
α	159.5	46	205.5	.62
β	581.25	209.25	790.5	.51
γ	936.75	396.75	1333.5	.43
δ	1095.5	496.5	1592	.40
ε	1136	545	1681	.37

C) CUMULATIONS BELOW CUTS

α	988.5	514	1502.5	.33
β	566.75	350.75	317.5	.22
γ	211.25	163.25	374.5	.13
δ	52.5	63.5	116	-.09
ε	12	15	27	-.11

D) ff-LOGS -- in matched form

Cut	f-log above	f-log below	Difference = ff-log
α	.62	.33	.29
β	.57	.22	.35
γ	.43	.13	.30
δ	.40	-.09	.49
ε	.37	-.11	.48

exhibit 2 (cont'd.)

E) NATURE OF TREND -- ??

Cut	Above* %	Above* f-log	ff-log	Straight - line?	Residual?
α	12%	-1.00	.29	.25	.04
β	46.4%	-.07	.35	.33	.02
γ	78%	.63	.30	.39	-.09
δ	93.2%	1.29	.49	.44	.05
ε	98.4%	2.06	.48	.50	-.02

Notes: * for athletic and non-athletic combined.
Straight-line? = by eye from plot of ff-log against f-log.
Residual? = ff-log minus straight-line? .

F) PROBLEM

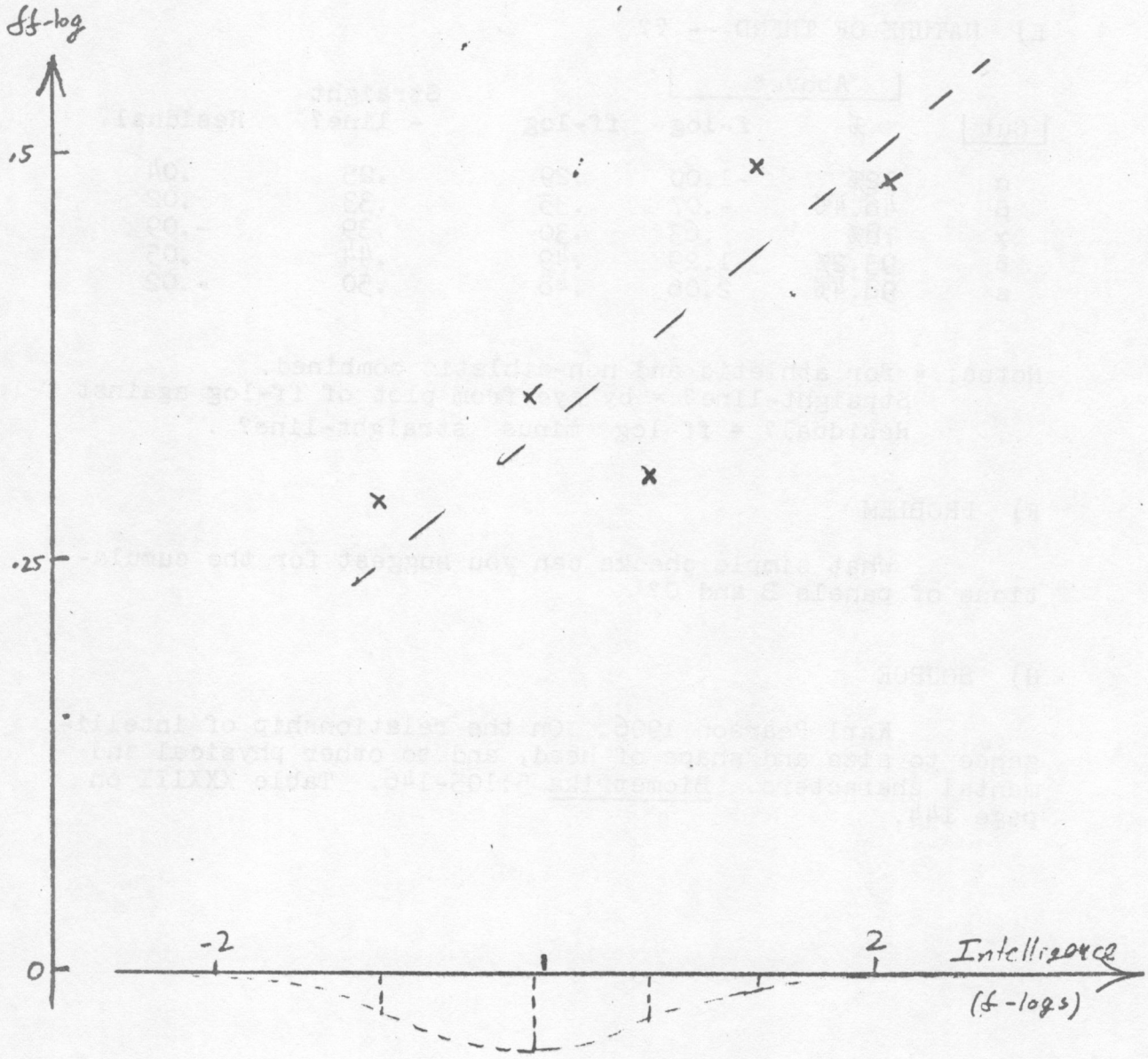
What simple checks can you suggest for the cumulations of panels B and C?

G) SOURCE

Karl Pearson 1906. On the relationship of intelligence to size and shape of head, and to other physical and mental characters. Biometrika 5:105-146. Table XXXIII on page 144.

exhibit 3 of chapter 29x -- school boys

Strength of relationship plotted against f-log of cutting point on intelligence



The example we have just seen was still somewhat special: One factor had only two versions. Exhibit 4, drawn from the same source, shows us the workings of a 6-by-6 case. The various double-flogs are in general agreement but considerable structure is clearly present

* some are weaker *

Before plunging ahead, we ought to pay attention to which double flogs are well determined and which are not. The two 2-by-2 tables:

$$\begin{array}{c|cc} & \multicolumn{2}{c}{\mu} \\ \beta & 477 & 409 \\ & 306 & 609 \end{array} \qquad \begin{array}{c|cc} & \multicolumn{2}{c}{\lambda} \\ \varepsilon & 125 & 1690 \\ & 1 & 35 \end{array}$$

correspond to matched ff-logs of .43 and .46. We would, I believe, all agree that the latter is much more "teetery" than the former. Moving any single school girl in the former cannot alter the double flog by as much as ±.01. (If in doubt, try a few moves.) Moving one school girl into or out of the SW cell in the latter has much larger effects, reducing the double flog by .30 on a move in -- and making it infinite on a move out.

These simple calculations somewhat overestimate the difference in stability between the two double flogs, but they point in the right direction. Thus, exhibit 5 begins (panel A)

exhibit 4 of chapter 29 -- school girls

Intellectual power and handwriting for school girls

A) The COUNTS

Intelligence		Handwriting						
Grade	Cut	Very good λ	Good μ	Moderate ν	Poor π	Bad ρ	Very bad	(sum
Quick Intelligent	α	38	115	67	26h	7	2	(255h)
Intelligent	β	47	277	231h	62	11	2	(630h)
Slow Intelligent	γ	30	185	244h	75h	9	4	(548)
Slow	δ	6	55h	112h	48	10	1	(233)
Slow Dull	ε	4	19	43	21	7	4	(98)
Very Dull		1	6	14	11	1	3	(36)
um		(126)	(657h)	(712h)	(244)	(45)	(16)	(1801)

B) CUMULATED COUNTS

Intelligence above	Handwriting above					
	λ	μ	ν	π	ρ	(All)
α	38	153	220	246h	253h	(255h)
β	85	477	775h	864	882	(886)
γ	115	692	1235	1399	1426	(1434)
δ	121	753h	1409	1621	1658	(1667)
ε	125	776h	1475	1708	1752	(1765)
(All)	(126)	(783h)	(1496)	(1740)	(1785)	(1801)

Examples: 38 = 38, 85 = 47 + 38, 115 = 30 + 85, ---

115 = 115, 392 = 277 + 115, 577 = 185 + 392, ---

153 = 115 + 38, 477 = 392 + 85, 692 = 577 + 115, ---

exhibit 4 (cont'd.)

C) INVERTED CUMULATIONS

Intelligence below	Handwriting above λ	μ	ν	π	ρ	(All)
(All)	(126)	(783h)	(1496)	(1740)	(1785)	(1801)
α	88	630h	1276	1493h	1531h	(1545h)
β	41	306h	720h	876	903	(915)
γ	11	91h	261	341	359	(367)
δ	5	30	87	119	127	(134)
ε	1	7	21	32	33	(36)

Examples: 1 = 1, 5 = 4 + 1, 11 = 6 + 5, ---
6 = 6, 25 = 19 + 6, 80.5 = 55.5 + 25, ---
7 = 6 + 1, 30 = 25 + 5, 91.5 = 80.5 + 11, ---
14 = 14, 57 = 43 + 14, 169.5 = 112.5 + 57, ---

D) % for CUMULATED COUNTS

Intelligence		For handwriting above λ	μ	ν	π	ρ
above	α	14.9%	60.0%	86.1%	97.3%	99.2%
	β	9.7	54.5	87.5	97.5	99.6h
	γ	8.0	48.2	86.2	97.6	99.4h
	δ	7.2	45.2	84.5	97.2	99.4h
	ε	7.0	44.0	83.6	96.8	99.2h
below	α	5.7%	40.8%	82.5%	96.6%	99.1%
	β	4.5	33.5	78.8	95.7	98.7
	γ	3.0	24.9	71.0	92.9	97.8
	δ	3.7	22.4	65.0	88.8	94.8
	ε	2.8	19.5	58.3	88.9	91.7

Examples: 14.9% = 38/255.5, 60.0 = 153/255.5, ---
5.7% = 88/1545.5, ---

exhibit 4 (cont'd.)

E) f-LOGS

Intelligence		For handwriting above λ	μ	ν	π	ρ
above	α	-.86	.20	.91	1.79	2.41
	β	-1.12	.09	.98	1.83	2.85
	γ	-1.22	-.04	.92	1.85	2.60
	δ	-1.28	-.10	.84	1.77	2.60
	ε	-1.29	-.12	.81	1.70	2.45
below	α	-1.40	-.18	.78	1.67	2.36
	β	-1.53	-.34	.65	1.55	2.17
	γ	-1.74	-.55	.45	1.28	1.90
	δ	-1.62	-.62	.31	1.04	1.45
	ε	-1.77	-.70	.17	1.04	1.20

F) ff-LOGS -- matched

Intelligence cut	Handwriting cut λ	μ	ν	π	ρ
α	.54	.38	.13	.12	.05
β	.41	.43	.33	.28	.68
γ	.52	.51	.47	.57	.70
δ	.34	.52	.53	.73	1.15
ε	.46	.58	.64	.66	1.25

Hinges of batch: .38 and .64. Centrality: = 2 = 51h/26. Outs: 2/25.

G) A TWO-WAY ANALYSIS -- three half-steps of median polish, unit = 0.01 in matched ff-log (or f-log).

	λ	μ	ν	π	ρ	eff	(f-log)
α	39	21	0	-11	-37	-34	(-90)
β	0	0	-6	-21	0	-8	(-2)
γ	3	0	0	0	-6	0	(68)
δ	-21	-5	0	10	33	6	(126)
ε	-14	-6	6	-2	38	11	(195)
eff	-2	0	-4	6	19	51	
(f-log)	(-129)	(-13)	(90)	(167)	(236)		

Hinges of residuals: -6 and +6. Clarity of fit: 6h = 68/12. Outs: 7/25.

exhibit 4 (cont'd.)

H) The FIT corresponding to panel G

15	17	13	23	42
41	43	39	49	68
49	51	47	57	76
55	57	53	63	82
60	64	58	68	87

I) SOURCE

Karl Pearson 1906. On the relationship of intelligence to size and shape of head, and to other physical and mental characters. Biometrika 5: 105-146. Table on page .

J) PROBLEMS

1) Compare panels F and G to check panel H.

2) Check that the fit of panel H is possible.

exhibit 5 of chapter 29x -- school girls

Looking at the firmer ff-logs of exhibit 4

A) COUNTS of SMALLEST CORNER -- of 2-by-2 table

	λ	μ	ν	π	ρ
α	38	102	35h	9h	2h
β	41	306h	101h	22	4
γ	11	91h	106	26	8
δ	5	30	47	15	7
ε	1	7	15	4	3

Examples: 38 from $\frac{38 \mid 217h}{88 \mid 1457h}$, 41 from $\frac{85 \mid 801}{41 \mid 874}$.

(2-by-2 tables can be found from panels B and C of exhibit 4)

B) TWO-WAY ANALYSIS of ff-LOGS -- after dropping row ε and column ρ, see panel F of exhibit 4 -- unit = 0.01

	λ	μ	ν	π	eff
α	26	9	-9	-12	-18
β	1	2	-1	-8	-6
γ	0	-2	1	9	6
δ	-21	-4	4	22	9
eff	2	3	-4	-2	44

Hinges of residuals: -6 and 6h. Outs: 3/16.

C) TWO-WAY ANALYSIS with ONE further ff-LOG treated as ABSENT -- unit = 0.01.

	λ	μ	ν	π	eff	(f-log)
α	25	9	9	-12	-18	(-90)
β	0	2	-1	-8	-6	(-2)
γ	-1	-2	1	9	6	(68)
δ	abs	-8	0	18	13	(126)
eff	3	3	-4	-2	44	
(f-log)	(-129)	(-13)	(-90)	(-167)		

Note: Residual corresponding to observed double flog treated as absent = -26.

exhibit 5 (cont'd.)

D) PROBLEM

1) Compare the residuals of panels B and C and show that their difference is a possible columns-plus-rows fit.

with a table of the smallest corners for the 2-by-2 tables of exhibit 4. If we want to drop rows and/or columns so as to raise the lowest of these lowest corner counts, the natural sequence of drops is:

first drop ε, raising lowest to 4 (highest dropped = 15)

then drop ρ, raising lowest to 5 (highest dropped = 8)

then drop δ, raising lowest to 9h (highest dropped = 47).

A plausible compromise is to drop ε and ρ, keeping δ since all but one of the counts is at least 15.

Panel B shows the two-way analysis of the resulting 4-by-4 table of double flogs. The situation is how much clearer. If we were afraid of the least corner count of 5 for column λ and row δ, we could treat this as an absent value as in panel C.

Either panel B or panel C tells the same story:

- small, presumably trivial effects for λ, μ, ν, and π.

- large, quite systematic effects for α, β, γ, δ.

- large, quite systematically arranged residuals.

Either of these panels shouts for further analysis.

29xC. Extending the analysis

The pattern of residuals in panels B and C of exhibit 5 is familiar (see chapter 18). By now, large positive values for one pair of diagonally opposite corners and large negative values for the other pair should scream at us, calling for re-expression or extension of the fit. Here it seems natural to try to extend the fit.

When we extended analyses before, we used what we could to construct a product carrier -- all we had then were the effects found by the columns-plus-rows fitting that defined the residuals. Now we are better off -- and should act accordingly. Not only do we have fitted effects -- which for columns at least are dangerously weak in our example -- but we have the f-logs themselves, which are rather clearly quite strong and solid.

If we are to construct a product carrier, then, it would seem that we ought to use the single flogs.

* a flog product carrier *

Using the single flogs to construct a product carrier introduces us to a new consideration

centering

When we used row and column effects for such a carrier, they were centered, in the sense that we had moved to the common term just enough to leave each set of effects "centered around

zero" in some specific sense. (If we analyzed by medians, we made the medians zero -- if by means, we made the means zero.) When we calculate flogs for each cutting point between columns, for example, it is not natural to center the set of flogs. (this is clearest when there are only two columns and one flog.)

We have, apparently, three choices:

- we can multiply together the flogs for rows and columns, and try to use them as is.
- we can multiply together the flogs, make a rows-plus-columns fit and use the residuals.
- we can center the two sets of flogs before multiplying them together.

Exhibit 6 shows -- at least for median analysis of our school-boy example -- that the last two alternatives are essentially the same -- the only differences are due to rounding in the arithmetic.

When we come to think about the mechanics of fitting, we find the first of our presumed alternatives unacceptable. We are extending a fit beyond columns-plus-rows fitting -- it is unsound to try to do this by fitting directly a carrier that overlaps columns or rows or both. Only one distinct alternative is left, given in two equivalent forms. Exhibit 7 shows the mechanics and the result. The reduction in midspread of residuals from 13h to 3h is striking evidence of the effectiveness of the fit.

exhibit 6 of chapter 29x -- school girls

Three versions of a flog-by-flog product carrier

A) The ROW and COLUMN FLOGS -- see panel A of exhibit 4 for counts, unit = 0.01 in raw flog

cut	λ	μ	ν	π	ρ
flog	-112	-11	69	145	205

cut	flog
α	-78
β	-1
γ	59
δ	110
ε	169

Examples: -112 from -1.12, -1.12 = log 126/1675;
-11 from -0.11, -0.11 = log 783h/1017h.

B) The PRODUCT CARRIER, dropping ε and ρ -- unit 0.01 for product

	-112	-11	69	145
-78	88	9	-54	-113
-1	1	0	-1	-1
59	-66	-7	41	86
110	-123	-12	76	160

Examples: 88 from .88, .88 = -1.12 times -.78;
1 from .01, .01 = -1.12 times -.01.
(In practice, 88 = one one-hundredth of -112 times -78, etc.)

C) COLUMN-plus-ROW ANALYSIS by MEDIANS of panel B -- unit 0.01

150	43	-44	-125	-30
42	12	-13	-35	-8
-42	-12	12	35	9
-114	-33	32	94	26
-51	-12	12	34	8

(cont'd.)

exhibit 6 (cont'd.)

D) MEDIAN-CENTERED VALUES of FLOGS and their PRODUCTS -- unit 0.01 throughout

	-141	-40	40	116
-104	151	43	-43	-124
-30	42	12	-12	-35
30	-42	-12	12	35
81	-114	-32	32	94

Examples: -141 = -112 - (29), -40 = -11 - (29), etc. where 29 is the median of -112, -11, 69, 145.

-107 = -78 - (29), etc., where this 29, a coincidence, is the median of -78, -1, 59, 110.

151 - (-141 times -107)/100, etc.

exhibit 7 of chapter 29x -- school girls

Extending the fit

A) The END THIRDS -- carrier from panel C of exhibit 6

	Low values of product carrier	Associated residuals in panel B of exhibit 5
	-125	-12
	-114	-21
	-44	-9
	-42	0
	-35	-8
(medians)	(-44)	(-9)

	High values of product carrier	
	150	26
	94	22
	43	9
	42	1
	35	9
(median)	(43)	(9)

$$\text{Tilt} = \frac{9 - (-9)}{43 - (-44)} = .21$$

B) The ADDITIONAL FIT 0.21 times panel C of exhibit 6, (0.21 times effects in parentheses)

32	9	-9	-26	(-6)
9	3	-3	-7	(-2)
-9	-3	3	7	(2)
-24	-7	7	20	(5)
(-9)	(-3)	(3)	(7)	(2)

Examples: 32 = .21 times 150,
9 = .21 times 42, etc.

(cont'd.)

exhibit 7 (cont'd.)

C) The NEW RESIDUALS -- panel B of exhibit 5 MINUS panel B above.

-6	0	0	14
-8	-1	2	-1
9	1	-2	2
3	3	-3	2

Examples: -6 = 26 - (32),
-8 = 1 - (9), etc.

D) TWO-WAY ANALYSIS of the NEW RESIDUALS -- an INNER ANALYSIS to that of panel B of exhibit 5 -- unit still 0.01

					eff
	-4	-1	1	12	0
	-4	0	5	-1	-2
	11	0	-1	0	0
	4	1	-3	-1	1
eff	-2	1	-1	2	0

Hinges of residuals: -1 and 2h. Outs: 2/16. Dets: 2/16.

Note that the two detached residuals correspond to smallest corners of only 11 and 9h, respectively.

E) The COLUMN-plus-ROW FIT of the COMBINED ANALYSIS

	λ	μ	ν	π
B of exh. 6	2	3	-4	-2
D above	-2	1	-1	2
sum	0	4	-5	0

	Panels G	Panels D	sum
α	-18	0	-18
β	-6	-2	-8
γ	6	0	6
δ	9	1	10

Common term: 44 + 0 = 44

* looking harder at our fit *

Two of the 16 residuals are detached. As noted below panel D, they are associated with small counts in the smallest corner -- actually with two of the three smallest. Since such residuals are expected to be more variable, we can relax about their detached condition.

When we look at the effects of the final (combined) column-plus-row analysis, we see results roughly as before:

- larger, systematically changing effects of row (intelligence).
- smaller irregular effects of column (handwriting).

One experiment we might like to try, then, is the result of fitting only by rows in addition to the product carrier. The simplest fit of this sort is found by adding back column effects to residuals and the common term to row effects. Panel A of exhibit 8 shows the result. The midspread is now intermediate in size between those of the previous fits. Panel B shows how much the extension improved the columns-plus-rows fit, and confirms that:

- rows alone extended is considerably better than columns-plus-rows.
- columns-plus-rows extended is considerably better still.

exhibit 8 of chapter 29x -- school girls

A less complete analysis and a comparison of analyses

A) The RESULT of ADDING BACK

	λ	μ	ν	π	eff		
α	-4	3	-4	12	26	AND	0.21 times panel C of exhibit 6
β	-4	4	0	-1	36		
γ	11	4	-6	0	50		
δ	2	3	-10	1	56		

Hinges of residuals: -4 and 3h. Outs: 1/16

B) COMPARISON of VARIOUS ANALYSES

Panel and Exhibit	Nature of fit	Constants fitted	Degrees of freedom	Midspread of Residuals	$\frac{(\text{Midspread})^2}{\text{Degrees of Freedom}}$
(F of 4)	Trimean only	1	15	19	24
B of 5	+ cols + rows	7	9	13h	20
D of 7	same extended	8	8	3h	1.5
A of 8	rows extended	5	11	7h	5.2

C) PROBLEM

1) Start with the 16 double flogs analyzed in panel B of exhibit 5, make a rows-only analysis, find the corresponding product carrier, extend the analysis, and compare the result with panel A above. How much improvement can we see?

How, then, shall we describe our extended columns-plus-rows fit? Exhibit 9 explores some alternatives. Panel A shows, each in two ways, the extension of the columns-plus-rows fit, and the columns-plus-rows fit. To use this we have to use an extension of the form

(constant)(column flog -.29)(row flog -.29)

where both shifts of .29 are special to this body of data. When we do this the row effects are quite systematic, differing from .16(row flog -.29) by no more than ±.02; while the column effects are irregular by ±.05.

Panel B shows that part of the fit which is of a column-plus-row character but is naturally associated with the extension. It amounts, of course, to

0.21{.29(row flog) + .29(col flog) - (.29)(.29)}.

If we transfer it, we get the breakdown in panel C, where the extension is taken in the form

(constant)(column flog)(row flog)

which is no longer special to this set of data. The row effects are now rather smaller, still systematically trending, but now more obviously flattening out at one end. They must still be within ±.02 of a multiple of (row flog -.29). This multiple is,

exhibit 9 of chapter 29x -- school girls

Some alternate expressions for the fit
yielding the residuals of panel D of exhibit 7

A) The EXTENSION described as FITTED -- unit 0.01

extension (as in panel B of exhibit 7)

numbers				formula
32	9	-9	-26	
9	3	-3	-7	0.21(col flog-.29)(row flog-.29)
-9	-3	3	7	
-24	-7	7	20	

columns-plus-rows (as in panel E of exhibit 7)

numbers				parts				
26	30	21	26	0	0	0	0	-17
36	40	31	36	0	0	0	0	-7
50	54	45	50	0	0	0	0	7
54	58	49	54	0	0	0	0	11
				0	4	-5	0	43

B) The COLUMNS-plus-ROW PART that comes NATURALLY with the EXTENSION -- from the parentheses of panel B of exhibit 7.

numbers				parts				
-13	-7	-1	3	0	0	0	0	-6
-9	-3	3	7	0	0	0	0	-2
-5	1	7	11	0	0	0	0	2
-2	4	10	14	0	0	0	0	5
				-9	-3	3	7	2

(cont'd.)

exhibit 9 (cont'd.)

C) The EXTENSION described MORE GENERALLY -- found by adding and subtracting panel B, unit = 0.01

extension (combined from panel B of exhibit 7

numbers				formula
19	2	-10	-23	
0	0	0	0	
-14	-2	10	18	0.21(col flog)(row flog)
-26	-3	17	34	

additional columns-plus-row fit

numbers				parts				
39	37	22	23	0	0	0	0	-11
45	43	28	29	0	0	0	0	-5
55	53	38	39	0	0	0	0	5
56	54	39	40	0	0	0	0	6
				9	7	-8	-7	41

D) The FIT produced by either panel A or panel C

fit				residuals			
58	39	12	0	-4	7	1	12
45	43	28	29	-4	9	5	-1
41	51	48	47	11	0	-1	0
30	51	56	74	4	1	-3	-1

Hinges of residuals: -1 and 2h. Clarity: 21 = (74-01/3h).
Centrality: 12 = 41/3h

E) PROBLEM

1) Given the analysis of problem 1 of exhibit 8, prepare the analogs of panels A and C above.

as can be seen by a short calculation, related to the multiple of .16 mentioned in describing panel A by .10 = .16 - (.29)(.21), The column effects now look quite different, being close to +.08 for the left two columns (better than moderate handwriting) and close to -.08 for the right two (poorer than moderate handwriting).

Either the panel A description or the panel C description does an equally good job of describing the original table of double flogs. To choose between them -- or to report both -- is a matter for the analyst's judgment.

29xD. <u>The smallest corners again -- 3 examples</u>

As we shall soon see by example, plotting the various double flogs that can be found for a doubly ordered contingency table -- one with each margin ordered -- against the log of the smallest corner often helps us to describe what is going on. It often pays to do this either BEFORE or IN PLACE OF fitting columns-plus-rows.

* men's vision *

Exhibit 10 sets out data -- used as an illustration by Kendall and Stuart -- on the quality of distance vision for 3242 male workers in World War II British munitions factories, and calculates the double flogs and log smallest corners and extracts the tilt of the former on the latter. The resulting fit of

exhibit 10 of chapter 29x -- men's vision

Unaided distance vision in the two eyes of 3242 men aged 30-39 employed in U.K. Royal Ordnance factories 1943-46

A) The COUNTS

Vision of right eye		Vision of left eye: Highest α	Second β	Third γ	Lowest	(sum)
Highest	α	821	112	85	35	(1053)
Second	β	116	494	145	27	(782)
Third	γ	72	151	583	87	(893)
Lowest		43	34	106	331	(514)
(sum)		(1052)	(791)	(919)	(480)	(3242)

B) The SMALLEST CORNERS

	counts α	counts β	counts γ	logs α	logs β	logs γ
α	231	120	35	2.36	2.08	1.54
β	115	292	62	2.06	2.47	1.79
γ	43	77	149	1.63	1.89	2.26

C) The DOUBLE and SINGLE FLOGS -- unit = 0.01

	α	β	γ	flog
α	147	109	70	-32
β	106	128	108	11
γ	80	102	150	73
flog	-32	12	76	

(cont'd.)

exhibit 10 (cont'd.)

D) FITTING tilt on LOG SMALLEST CORNERS -- medians bold

Third	log smallest corner	double flog	.92 times log s c	Diff	less -.72
Low	1.54	.70	1.41	-.71	.01
	1.63	**.80**	1.50	-.70	.02
	1.79	1.08	1.64	-.56	.16
Mid	1.89	1.02	1.74	-.72	.00
	2.06	1.06	1.89	-.83	-.11
	2.08	1.09	1.91	-.82	-.10
High	2.26	1.50	2.08	-.58	.14
	2.36	**1.47**	2.16	-.69	.03
	2.47	1.28	2.26	-.98	-.26

$$.92 = \frac{1.47 - .80}{2.36 - 1.63}$$

E) RESIDUALS in 2-WAY FORM -- unit 0.01

	α	β	γ
α	3	-10	1
β	-11	-26	16
γ	2	0	14

F) SOURCE

M. G. Kendall and A. Stuart 1961. The Advanced Theory of Statistics (3-volume edition). Table 33.3 at page 564 of volume 2.

-.72 plus .92 log smallest corner

leaves residuals ranging from -.26 to +.16 in a pattern shown in panel E. The grouping together of substantially negative and substantially positive residuals is suggestive but probably not convincing. As exhibit 11 shows, fitting a tilt against log smallest corner is clearly called for.

* call money *

Exhibit 12 sets out data on the New York City money market for about 15 years in the last century. Panels D and E show graphically -- in coded form -- the patterns of the double flogs and the logs of smallest corners. Not much resemblance is evident, though concentrations of low values around the NE margins show in both. Exhibit 13 plots the eight summary points derived in problem 2 of exhibit 12. Only the point for smallest count = 1 seems at all distinctive. No trend is apparent, particularly when we discount this point a little.

* bean dimensions *

Exhibit 14 opens with a distribution of lengths and breadths for 9440 beans. These are then converted into double flogs and log smallest counts and displayed in coded form. Clearly the patterns tend to be opposite, clearly indicating a negative tilt of double flog against log smallest corner.

Plot of the nine summary points and group-fitted line

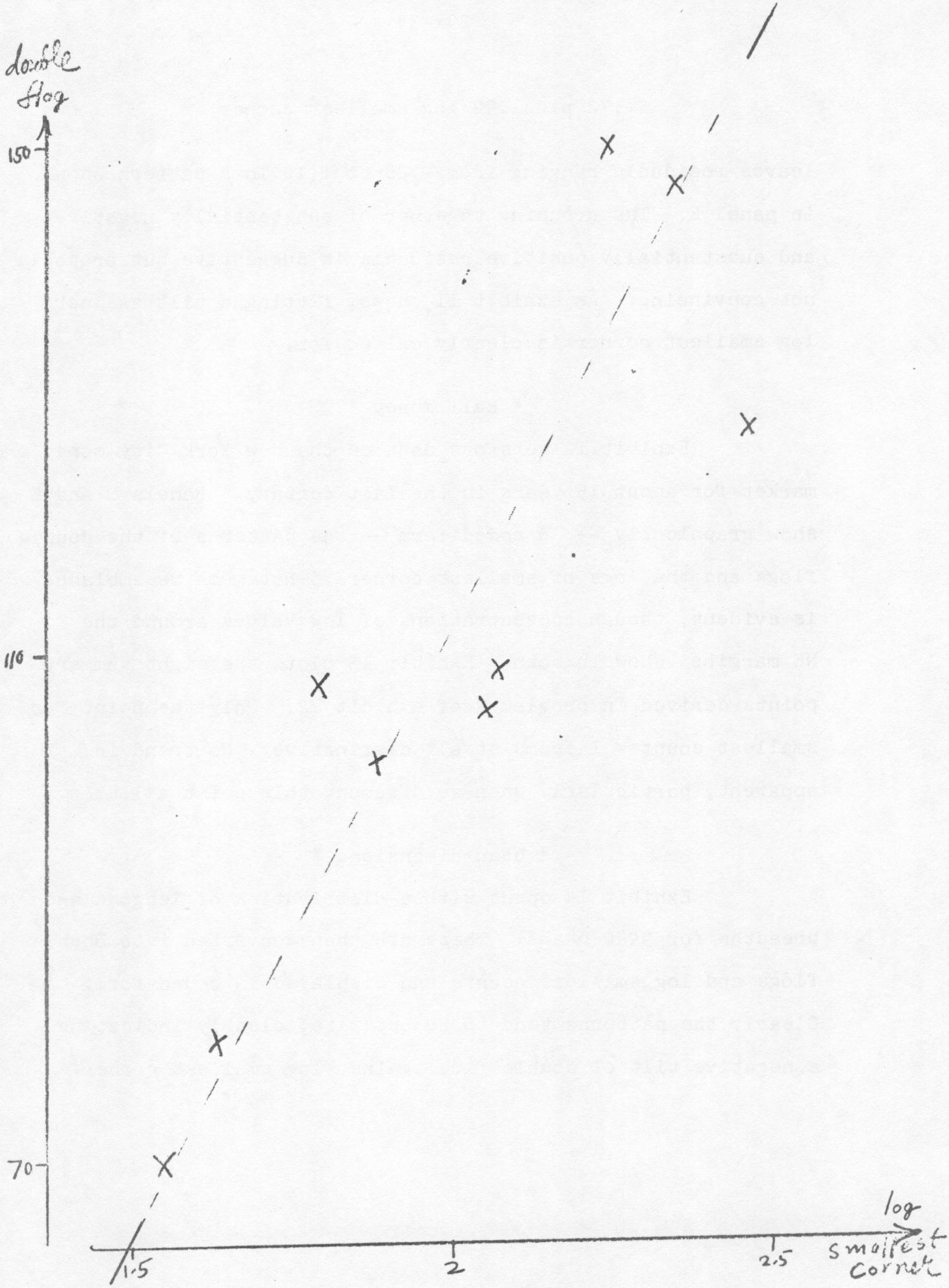


exhibit 12 of chapter 29x -- call money

New York call discount rates and reserves in New York associated banks (as percentages of deposits)

A) The COUNTS -- condensed as indicated, zeros not entered

% Reserve	Call discount rates in %											
	1	a 1.5 2	b 2.5 3	c 3.5 4	d 4.5 5	e 5.5 6	f 6.5 7	g 8 9	h 10 12	i 15 20	j 25	(Total)
21-23	-	-	-	1	2	-	-	1	-	-	-	(4)
p 24-25	-	-	-	3	11	17	3	10	5	1	1	(51)
q 26-27	-	11	25	46	64	35	18	3	2	3	2	(209)
r 28-29	3	92	76	27	14	3	7		1		1	(224)
s 30-31	9	40	24	12	4							(89)
t 32-33	30	36	14	3	2							(85)
u 34-35	10	18										(28)
v 36-37	15	4										(19)
w 38-39	28	4										(32)
x 40 up	26	13										(39)
(Total)	(121)	(218)	(139)	(92)	(97)	(55)	(28)	(14)	(8)	(4)	(4)	(780)

B) The SMALLEST CORNERS -- zeros not entered

	a	b	c	d	e	f	g	h	i	j
p	-	-	-	1	1	1	1	-	-	-
q	-	-	-	4	17	21	10	7	2	1
r	-	11	36	32	12	9	2	2	1	1
s	3	59	21	6						
t	12	19	5	2						
u	21									
v	36									
w	13									
x	2									

(cont'd.)

exhibit 12 (cont'd.)

C) The DOUBLE FLOGS -- unit 0.01 raw

	a	b	c	d	e	f	g	h	i	j	(flog)
p				93	30	62	93				(229)
q				166	128	106	146	106	66	65	(112)
r		160	158	143	142	111	148	117	115	67	(29)
s	204	115	124	153							(22)
t	174	142	162	174							(-45)
u	160										(-75)
v	116										(-88)
w	113										(-100)
x	156										(-128)
(flog)	(-74)	(-11)	(20)	(43)	(77)	(110)	(140)	(168)	(198)	(229)	

Median: 126. Hinges: 106 and 157. 8ths: 66 and 164.

D) CODED DOUBLE FLOGS

p				o	0	o	o			
q				x	.	.	.	.	o	o
r		x	x	.	.	.	.	.	.	o
s	x	.	.	.						
t	x	.	x	x						
u	x									
v	.									
w	.									
x	x									

(cont'd.)

29x-37

exhibit 12 (cont'd.)

E) CODED LOG SMALLEST CORNERS

	a	b	c	d	e	f	g	h	i	j
p				o	o	o				
q				.	.	x	.	.	o	
r		.	x	x	.	.	.	o	o	
s	.	x	x	.						
t	.	x	.	.						
u	x									
v	x									
w	.									
x	.									

F) SOURCE

M. G. Kendall 1943. The Advanced Theory of Statistics, volume 1. Table 1.26 on page 28.
His source:
J. P. Norton 1902. Statistical studies in the New York Money Market. Macmillan.

G) PROBLEMS

1) Plot double flogs against log smallest corners. What can you see?

2) Show that when the log smallest corners are divided into groups of 3 or more, preserving order and ties the following pairs of medians can be found -- pattern is (log smallest corner, double flog) -- (00,67), (30,148), (65,164), (95,111), (108,160), (123,128), (132,124), (156,130) as plotted in exhibit 13.

29x-38

The 8 summary points of problem 2 of exhibit 12

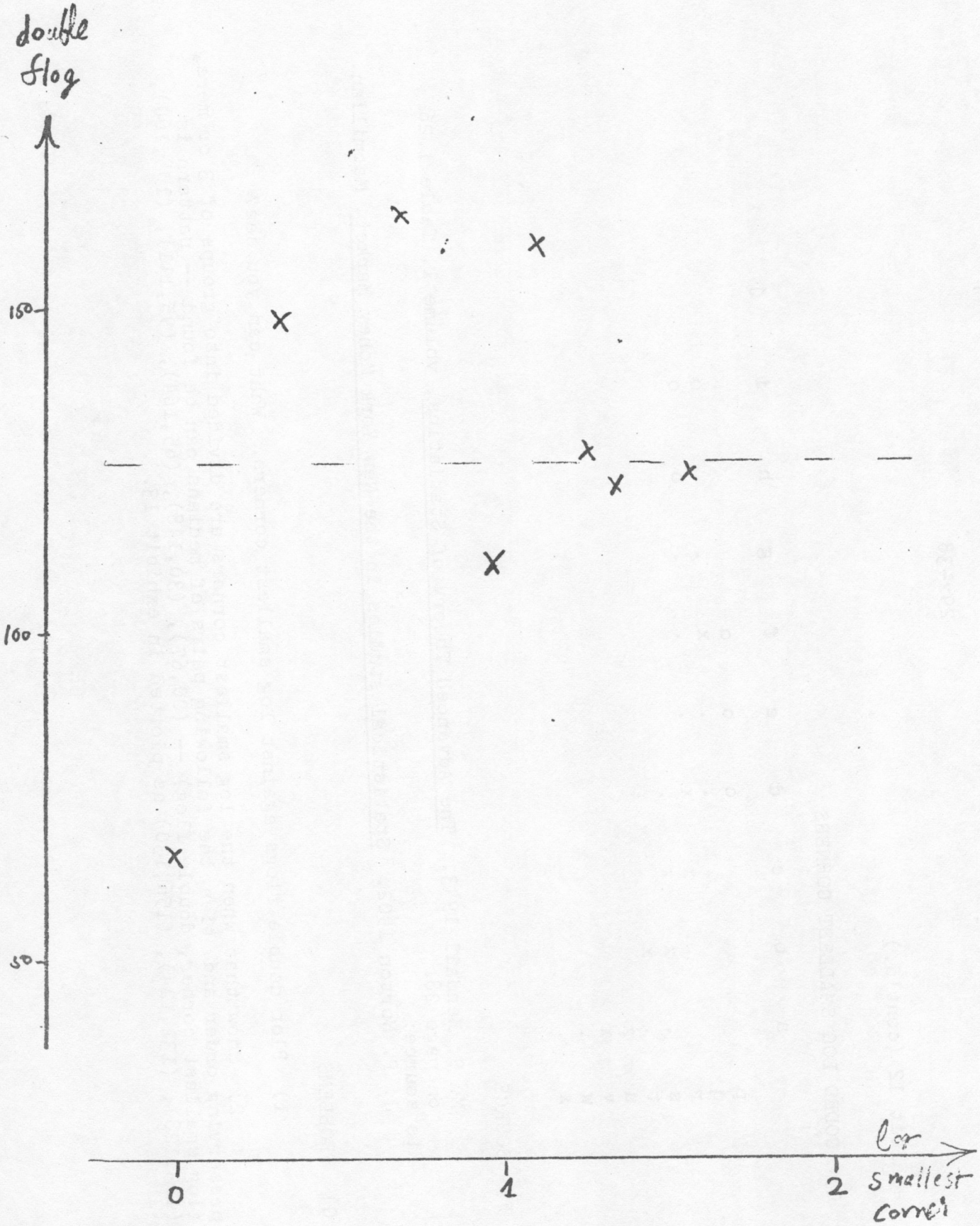


exhibit 14 of chapter 29x -- bean dimensions

Lengths and breadths of 9440 beans

A) The COUNTS

breadths*	lengths (center of interval, unit 0.25 millimeter)															
	68	66	64	62	60	58	56	54	52	50	48	46	44	42	40	38
73	-	2	-	-	3	-										
71	4	8	17	19	-	-	-									
69	2	23	101	156	93	23	2	-								
67	-	18	105	494	574	227	56	9	-	-						
65		4	44	375	956	913	362	73	12	3	-					
63		-	7	81	385	871	794	330	89	19	3	-				
61			1	4	65	236	469	361	175	55	27	4	-		-	
59			-	-	6	23	91	137	124	78	37	22	11	-	1	-
57						1	13	18	28	35	25	32	11	6	1	-
55						-	-	1	9	8	21	12	13	7	1	-
53								-	-	-	2	-	1	4	3	-
51										1	-	-	-	1	1	1

*unit: 0.125 millimeter

B) LOGS of SMALLEST CORNERS -- unit 0.01

cut for breadth*	cut for length (unit 0.25 millimeter)													
	67	65	64	61	59	57	55	53	51	49	47	45	43	41
72	-	30	30	30	-									
70	30	115	149	48	-	-								
68	-	134	220	208	140	30	-							
66	-	60	175	271	250	183	95	-	-					
64		-	90	197	274	271	199	118	48					
62		-	0	70	188	253	272	210	140	48	-			
60			-	-	78	148	213	246	205	153	60	-	-	-
58					-	0	115	152	184	202	158	108	0	0
56						-	-	0	100	128	162	148	90	30
54								-	-	0	48	48	60	48
52									-	0	0	0	00	30

*unit 0.125 millimeter

(cont'd.)

exhibit 14 (cont'd.)

Median: 112. Hinges: 30 and 184. 8ths: 0 and 216. 16ths: 0 and 262. Outs: 0/66

Note: Bold values are the largest, and are closest to col flog = 0 = row flog

C) DOUBLE FLOGS -- raw, unit 0.01

cut for breadth	cut for length (unit 0.25 millimeter)															(flog)
	67	65	63	61	59	57	55	53	51	49	47	45	43	41	39	
72	-	202	126	55	-											(-328)
70	198	184	154	197	-	-										(-225)
68	-	159	142	128	151	218	-									(-130)
66	-	176	135	111	117	139	183	-	-							(-59)
64		-	165	132	**114**	**117**	142	184	220	-						(-1)
62		-	172	159	140	121	124	148	184	248	-					(52)
60			-	-	196	171	149	156	176	198	262	-				(104)
58					-	264	192	194	197	203	214	226	300	240		(155)
56						-	-	296	227	225	205	227	249	259		(204)
54								-	-	271	245	283	317	324		(283)
52									-	206	234	250	309	320		(337)
(flog)	(-320)	(-218)	(-137)	(-70)	(-22)	(21)	(62)	(99)	(130)	(157)	(185)	(218)	(256)	(307)		

Median: 195. Hinges: 151h and 230h. 8ths: 127 and 267h. 16ths: 117 and 304h. Outs: 4/66.

Note: Two bold values are closest to center, namely to

col flog = 0 = row flog

(cont'd.)

exhibit 14 (cont'd.)

D) Coded DOUBLE FLOGS

```
o   .   o   O
.   .   .   .
    .   o   o   o   ,
    .   o   o   o   o   .
        .   o   o   o   o   .   .
        .   .   o   o   o   o   .   x
                .   .   o   .   .   .   x
                    x   .   .   .   .   .   x   x
                            x   .   .   .   x   x
                                    x   x   x   X   X
                                    .   x   x   x   X
```

E) Coded LOG SMALLEST CORNERS -- 30s included as adjacent low

```
-   o   o   o
o   .   .   .
    .   x   x   .   o
        .   x   x   .   .
        .   x   x   x   x   .
        o   .   x   x   x   x   .
                .   .   x   x   .   .
                    o   .   .   .   .   .
                        o   x   x   .   .   o   o
                            .   .   .   .   .   o
                                    o   .   .   .   .
                                    o   o   o   o   o
```

F) PROBLEM

Plot double flog against log smallest corner .

If we plot all 66 pairs, as suggested in the problem, we see -- did we not know already:

- that there is a negative tilt.
- that values are more variable when log smallest count is small.

We need to take out this tilt and see what is left.

Exhibit 15 applies the techniques of chapter 10 to the construction of 9 summary points. The tilt seems to be about .43. The median of

double flog PLUS .43 log smallest count

is about 2.40, leading to the residuals in panel C of exhibit 15. The midspread or hinge spread has been reduced from .79 to .50, the 8th spread from 1.40 to .85, and the 16th spread from 1.87 to 1.11. Clearly fitting a tilt against log smallest corner -- a negative tilt -- has made a great improvement in our partial description.

Equally clearly, as panel D shows, there is still much structure in our residuals. We ought to carry the fit further, as we will in a later section of this chapter.

* discussion *

We have seen three examples in this section, showing three kinds of relation of double flogs to log smallest count: positive tilt, no tilt, negative tilt. Further examples of all

exhibit 15 of chapter 29x -- bean dimensions

Finding 9 summary points, a tilt, and the resulting residuals

A) Calculations based on a stem-and-leaf for LOG SMALLEST CORNER -- unit = 0.01

|To|		log smallest corner	|median|	double flogs	|median|
10	0	0000000000	|0|	172,264,296,271,206,234,250,300,309,240	|257|
17	3	0000000	|30|	198,202,126,55,218,259,324,	|202|
23	4	888888	|108|	197,220,248,245,263,324,	|246|
26	6	000	|70|	176,262,317,	
28	7	08		159,196	|196|
31	9	0-0/5		165,-,249/183	
33	10	08	|112|	227,226,	
35	11	558		184,192,184	|188|
30	12	8		225,	
29	13	4	|144|	159,	|178|
28	14	90808		154,151,171,184,227,	
23	15	-38 \2		-,198,214, \194	
20	16	2		205	|197|
19	17	5	|175|	135	
18	18	834		140,139,197,	
15	19	79		132,142	
13	20	852	|206|	128,176,203	|145|
10	21	30		149,148,	
8	22	0		142,	
7	24	6		156,	
6	25	03	|271|	117,124	|117|
	26			.....	
4	27	1412		111,114,117,124,	

Means of end three: (26,235), (217,153)

Apparent tilt: $-.43 = \frac{153 - 235}{217 - 26}$

Note: "__" indicates value moved across dividing line.

(cont'd.)

exhibit 15 (cont'd.)

B) Calculation of DOUBLE FLOG plus .45 times LOG SMALLEST CORNER -- based on same format as panel A, unit = 0.01

	.43 log smallest corner	adjusted value = sum	median
0	all 0,	172,264,296,271,206,234,250,300,309,240,	257
3	all 13,	211,215,139,68,231,272,337,	215
4	all 21,	218,241,269,266,284,345	267
6	all 26,	202,288,343,	
7	30,34,	189,230	230
9	39,-,39/41	204,-,288/224	
10	43,46	270,272,	
11	49,51,51	233,243,235	243
12	55,	280,	
13	58,	217,	
14	64,60,64,60,64	218,211,235,244,291	
15	-,66,68 | 66	-,264,282 | 260	
16	70	275	
17	75	210	264
18	81,79,79,	221,218,276,	
19	85,86	217,228	
20	90,88,87,	218,264,290,	
21	92,90	241,238	238
22	95,	237,	
24	106,	262	
25	108,109,	225,233,	233
27	117,118,117,117,	228,232,234,241,	

(cont'd.)

exhibit 15 (cont'd.)

C) Display of RESIDUALS -- values of double flog PLUS .43(log smallest corner) MINUS 2.40
unit still 0.01

a	-25	-101	-172										
-29	-7	-22	-22										
	-23	-3	-22	-29	-9								
	-38	-30	-12	-14	-22	-9							
		-36	-23	-8	-6	-12	-5	2					
		-68	-51	-19	-7	1	-2	4	29				
				-10	-5	1	22	24	24	48			
					24	3	20	36	50	42	32	60	0
							56	30	40	35	51	48	32
									31	26	44	103	105
									-34	-6	10	69	97

Medians: -1. Hinges: -22 and +32. 8ths: -30 and +49. 16ths: -44 and +64. Outs: 5/66. Dets: 1/66

D) Display of CODED RESIDUALS

.	o	0	0										
o	.	.	.										
	o	.	.	o	.								
	o	o	.	.	0	.							
		o	o	.	.	.	.	.					
		o	o	.	.	.	.	.	.				
				.	.	.	.	.	.	x			
					.	.	.	o	o	x	.	x	.
							o	.	o	x	x	x	.
									.	.	x	X	X
									x	.	.	x	X

three are not hard to find. The conclusion to which we are thus pointed is simple: We ought to expect to use at least two constants, -- giving level AND tilt against log smallest corner -- to describe a table of double flogs.

In taking this position, we have given a substantial place to log smallest corner, to a quantity which first appeared as merely an indicator of relative variability. We can suggest some reasons for some of its importances -- and will do so shortly -- but our insight is quite limited. We cannot be said to understand why it works as well as it does.

29xE. Theoretical distributions

When we were looking at batches of numbers we found it convenient -- especially in connection with suspended rootograms -- to make use of standard patterns of distribution: the Gaussian, the Cauchy, the rectangular -- and even the rootistic, logistic, tooristic and invistic. Statisticians have been long concerned with such standard patterns, and have generated many individual patterns and several systems of patterns. Statisticians have varied widely in how seriously they have taken the details of such theoretical distributions and in how closely related to the behavior of actual data they have judged these details to be.

We are now concerned with patterns of distribution for (x,y) pairs. For a contingency table is just an account of

the pattern of distribution of

$$x = \text{column version}$$

$$y = \text{row version}$$

taken together. In those cases where the values of x are ordered, as are those of y, so that x and y can be reasonably assigned numerical values, whether or not they had such to begin with, we are essentially concerned with a pattern of distribution for pairs of numbers.

It is now natural to ask what standard patterns of distribution for pairs of numbers are either well-known or available. Only one system of patterns is at all well known. This is the system whose members are often labelled: "bivariate normal distributions" or "bivariate Gaussian distribution". For simplicity, we shall call them "biGaussian". These distributions are characterized by the fact that not only do x and y, separately, follow Gaussian distributions, but that the same is true for every $ax + by$ where a and b are fixed numbers, not both zero. Exhibits 16, 17, 18. 19 and 20 show examples of biGaussian patterns of distribution.

We have chosen to look at relationship in terms of tables of double flogs and their summarization by interesting fits. What do we expect from a biGaussian pattern of distribution? Let us go and look.

exhibit 16 of chapter 29x

Counts for a total of 1,000,000 fitting a biGaussian pattern
of distribution as closely as possible
No relationship

A) The COUNTS -- equispaced cuts

39	102	274	570	931	1189	1189	931	570	274	102	39	(6210)
102	275	720	1527	2479	3167	3167	2479	1527	720	275	102	(16540)
274	720	1951	4039	6603	8436	8436	6603	4039	1951	720	274	(44058)
570	1527	4039	8437	13766	17585	17585	13766	8437	4039	1527	570	(91848)
931	2479	6603	13766	22465	28698	28698	22465	13766	6603	2479	931	(149884)
1189	3167	8436	17585	28698	36655	36655	28698	17585	8436	3167	1189	(191460)
1189	3167	8436	17585	28698	36655	36655	28698	17585	8436	3167	1189	(191460)
931	2479	6603	13766	22465	28698	28698	22465	13766	6603	2479	931	(149884)
570	1527	4039	8437	13766	17585	17585	13766	8437	4039	1527	570	(91848)
274	720	1957	4029	6603	8436	8436	6603	4039	1957	720	274	(44058)
102	275	720	1527	2479	3167	3167	2479	1527	720	275	102	(18540)
39	102	278	570	931	1189	1189	931	570	274	102	39	(6210)

(sums) (6210) (44058) (149884) (191460) (91848) (18540) 1000000

(16540) (91848) (191460) (149884) (44058) (6210)

correlation coefficient = p = 0.0
coefficient of determination = p^2 = 0.00
Fisher's z = 1.15 log $\frac{1+p}{1-p}$ = 0.00

B) REMARK

All double flogs are zero.

exhibit 17 of chapter 29x

Counts of a total of 1,000,000 fitting a biGaussian pattern
of distribution as closely as possible
A case of weak relationship

A) The COUNTS -- equispaced cuts

												(total)
1	8	35	123	329	674	1058	1272	1172	828	448	262	(6230)
8	36	152	481	1175	2188	3120	3402	2837	1810	883	448	(16540)
35	152	581	1716	3863	6654	8765	8832	6608	4013	1810	828	(44057)
123	481	1716	4669	9715	15456	18831	17547	12484	6812	2837	1172	(91839)
329	1175	3863	9715	18688	27494	30940	26625	17547	8832	3402	1272	(149882)
674	2188	6654	15466	27494	37386	38886	30940	18831	8765	3120	1058	(191460)
1058	3120	8765	18831	30940	38886	37386	27494	15466	6654	2188	674	(191460)
1272	3402	8832	17547	26625	30940	27494	18688	9715	3863	1175	329	(149882)
1172	2837	6812	12484	17547	18831	15456	9715	4669	1716	481	123	(91839)
828	1810	4013	6808	8832	8765	6654	3863	1716	581	152	35	(44057)
448	883	1810	2837	3402	3120	2188	1175	481	152	36	8	(16540)
262	448	828	1172	1272	1058	674	329	123	35	8	1	(6230)

(sum) (6230) (44057) (149882) (191460) (91839) (16540)
(16540) (91839) (191460) (149882) (44057) (6230)

correlation coefficient = $\rho = 0.3$

coefficient of determination = $\rho^2 = 0.09$

Fisher's $z = 1.15 \log \frac{1+\rho}{1-\rho} = 0.31$

exhibit 18 of chapter 29x

Counts of a total of 1,000,000 fitting a biGaussian pattern
of distribution as closely as possible
A case of moderate relationship

A) The COUNTS -- equispaced cuts

.	.	.	3	19	96	331	786	1296	1485	1185	1009	(6210)
.	.	4	28	162	634	1705	3159	4021	3521	2121	1185	(16540)
.	4	33	220	1007	3164	6833	10145	10363	7282	3521	1485	(44057)
3	36	212	1166	4319	10863	18801	22406	18372	10363	4021	1296	(91848)
19	162	1007	4309	12663	25597	35588	34044	22406	10145	3159	786	(149883)
96	634	3164	10853	25597	41520	46540	35558	18801	6833	1705	331	(191662)
331	1705	6833	18801	35588	46540	41520	25597	10853	3164	634	96	(191662)
786	3159	10145	22406	34044	35588	25597	12663	4309	1007	162	19	(149883)
1296	4021	10303	18372	22406	18801	10863	4319	1166	212	36	3	(91848)
1485	3521	7282	10363	10145	6833	3164	1007	220	33	4	.	(44057)
1185	2121	3521	4021	3159	1705	634	162	28	4	.	.	(16540)
1009	1185	1485	1296	786	331	96	19	3	.	.	.	(6210)

(sum) (6210) (44057) (149883) (191662) (91848) (16540)
(16540) (91848) (191662) (149883) (44057) (6210)

correlation coefficient = $\rho = 0.6$

coefficient of determination = $\rho^2 = 0.36$

Fisher's $z = 1.15 \log \frac{1+\rho}{1-\rho} = 0.69h$

exhibit 19 of chapter 29x

Counts of a total of 1,000,000 fitting a biGaussian pattern
of distribution as closely as possible
A case of strong relationship

A) The COUNTS -- equispaced cuts

.	.	.	.	.	.	.	4	96	777	2115	3218	(6210)
.	.	.	.	.	.	7	236	2121	6147	5913	2115	(16539)
.	.	.	.	.	14	456	4037	15288	16799	6147	777	(44058)
.	.	.	.	18	673	7811	29308	36532	15288	2121	96	(91847)
.	.	.	18	767	10140	43309	61465	29308	4637	236	4	(149884)
.	.	14	673	10140	59793	79723	43309	7811	456	7	.	(191457)
.	7	456	7811	43309	79723	59793	10140	673	14	.	.	(191457)
4	236	4637	29308	61465	43309	10140	767	18	.	.	.	(149884)
96	2121	15288	36532	29308	7811	673	18	.	.	.	.	(91847)
2115	6147	16799	15268	4637	456	14	.	.	.	.	.	(44058)
2115	5913	6147	2421	236	7	.	.	.	.	.	.	(16539)
3218	2115	777	96	4	.	.	.	.	.	.	.	(6210)

(sum) (6210) (16539) (44058) (91847) (149884) (191457) (191457) (149884) (91847) (44058) (16539) (6210)

correlation coefficient = $\rho = 0.9$

coefficient of determination = $\rho^2 = 0.81$

Fisher's $z = 1.15 \log \frac{1+\rho}{1-\rho} = 1.47$

exhibit 20 of chapter 29x

Counts of a total of 1,000,000 fitting a biGaussian pattern
of distribution as closely as possible
A case of very strong relationship

A) The COUNTS -- equispaced cuts

.	.	.	.	.	.	.	.	4	231	1928	4047	(6210)
.	.	.	.	.	.	.	8	638	5845	8121	1928	(16560)
.	.	.	.	.	.	16	1372	14215	22378	5845	231	(44057)
.	.	.	.	.	24	2283	26822	47862	14215	638	4	(91849)
.	.	.	.	27	2946	39260	79448	26822	1372	8	.	(149884)
.	.	.	24	2946	44575	102362	39260	2283	16	.	.	(191460)
.	.	16	2283	39260	102362	44575	2946	24	.	.	.	(191460)
.	8	1372	26822	79448	39260	2946	27	.	.	.	.	(149884)
4	638	14215	47862	26822	2283	24	.	.	.	.	.	(91849)
231	5845	22378	14215	1372	16	.	.	.	.	.	.	(44057)
1928	8121	5845	638	8	.	.	.	.	.	.	.	(16560)
4047	1928	231	4	.	.	.	.	.	.	.	.	(6210)

(sum) (6210) (44057) (49884) (191460) (91849) (16560)
(16560) (91849) (191460) (149884) (44057) (6210)

correlation coefficient = ρ = .95

coefficient of determination = ρ^2 = .9025

Fisher's z = 1.15 log $\frac{1+\rho}{1-\rho}$ = 1.832

* double flogs for biGaussians *

We do not need to convert exhibit 16 into double flogs. We know that they are all zero. For the next four exhibits, this is not the case. Exhibits 21, 22, 23 and 24 show the tables of double flogs and log smallest corners that arise from exhibits 17, 18, 19 and 20 respectively. Exhibits 25 and 26 show the corresponding plots. We see that a straight-line fit does really quite well, though not perfectly.

* residuals after log smallest corner *

After making adjustments for both a tilt against log smallest corner and a constant, we reach the residuals shown in exhibit 27. We see that the residuals are pleasantly small tending to be positive where col flog = row flog -- in our exhibit along the diagonal from SW to NE -- to be negative at a short distance from this diagonal, and to be positive again still further away.

* deeper residuals *

Exhibit 28 shows these residuals plotted against

diff = column flog - row flog

for two of the cases -- the other two are left as problems. We now see the rather W-shaped behavior of the residuals as a function of diff rather more clearly.

exhibit 21 of chapter 29x

Table of double flogs and log smallest corners for
the biGaussian pattern of distribution shown in exhibit
Weak relationship -- ρ = .3

A) The DOUBLE FLOGS -- raw

											(f-log)
1.42	1.21	1.00	.84	.70	.64	.60	.62	.67	.76	.87	(2.20)
1.21	1.01	.84	.70	.62	.54	.52	.53	.60	.66	.76	(1.63)
1.00	.84	.70	.58	.50	.45	.44	.46	.52	.60	.67	(1.15)
.84	.70	.58	.49	.42	.39	.38	.41	.46	.53	.62	(.73)
.70	.62	.50	.42	.38	.36	.36	.38	.44	.52	.60	(.35)
.64	.54	.45	.39	.36	.34	.36	.39	.45	.54	.66	(.00)
.60	.52	.44	.38	.36	.36	.38	.42	.50	.62	.70	(-.35)
.62	.53	.46	.41	.38	.39	.42	.49	.58	.70	.84	(-.73)
.67	.60	.52	.46	.44	.45	.50	.58	.70	.84	1.00	(-1.15)
.76	.66	.60	.53	.52	.54	.62	.70	.84	1.01	1.21	(-1.63)
.87	.76	.67	.62	.60	.64	.70	.84	1.00	1.21	1.42	(-2.20)
(f-log) (-2.20)	(-1.63)	(-1.15)	(-.73)	(-.35)	(.00)	(.35)	(.73)	(1.15)	(1.63)	(2.20)	

B) The LOG SMALLEST CORNERS -- unit 0.01

0	95	164	222	270	307	334	343	319	285	242
95	172	238	293	337	372	397	394	367	331	285
164	238	300	352	394	426	449	435	405	367	319
222	293	352	401	440	470	488	466	435	394	343
270	337	394	440	477	505	512	688	449	397	334
307	372	426	410	505	**530**	505	470	426	372	307
334	397	449	488	512	505	477	440	394	337	207
343	394	435	466	488	470	440	401	352	293	222
319	367	405	435	449	426	394	352	300	238	164
285	331	367	394	397	372	337	293	238	172	95
242	285	319	343	336	307	270	222	164	95	0

C) PROBLEMS

1. Convert panels A and B to coded form. Compare.

2. Fit panel B with a straight line against $(\text{sum})^2 + (\text{coeff})(\text{diff})^2$ where

sum = (column flog) plus (row flog)
diff = (column flog) minus (row flog)

exhibit 22 of chapter 29x

Table of double flogs and log smallest corners for
the biGaussian pattern of distribution shown in exhibit
Moderate relationship $\rho = .6$

A) The DOUBLE FLOGS -- raw

											(flog)
.	.	.	2.59	2.29	1.72	1.37	1.25	1.33	1.41	1.56	(2.20)
.	.	2.62	2.30	1.68	1.38	1.19	1.12	1.15	1.25	1.41	(1.63)
.	2.62	2.11	1.67	1.33	1.11	.99	.96	1.02	1.15	1.33	(1.15)
2.59	2.30	1.67	1.32	1.07	.91	.85	.87	.96	1.12	1.25	(.73)
2.29	1.68	1.33	1.07	.89	.79	.78	85	.99	1.19	1.37	(.35)
1.72	1.38	1.11	.91	.79	**.76**	.79	.91	1.11	1.38	1.72	(.0)
1.37	1.19	.99	.85	.78	.79	.89	1.07	1.33	1.68	2.29	(-.35)
1.25	1.12	.96	.87	.85	.91	1.07	1.32	1.67	2.30	2.59	(-.73)
1.33	1.15	1.02	.96	.99	1.11	1.33	1.07	2.11	2.62	.	(-1.15)
1.41	1.25	1.15	1.12	1.19	1.38	1.68	2.30	2.62	.	.	(-1.63)
1.56	1.41	1.33	1.25	1.37	1.72	2.29	2.59	.	.	.	(-2.20)
(f-log)(-2.20)	(-1.63)	(-1.15)	(-.73)	(-.35)	(0)	(.35)	(.73)	(1.15)	(1.63)	(2.20)	

B) The LOG SMALLEST CORNERS -- unit = 0.01

.	.	.	48	134	207	265	309	340	334	300
.	.	60	154	233	298	347	384	402	374	334
.	60	161	247	317	373	415	445	436	402	340
48	154	247	323	376	434	469	486	445	384	309
134	233	317	326	441	482	511	469	415	347	265
207	298	373	434	482	**517**	482	434	373	298	207
265	347	415	469	511	482	441	376	317	233	134
309	384	445	496	469	434	376	323	247	154	48
340	402	436	465	415	373	317	247	161	60	.
334	374	402	347	347	298	233	154	60	.	.
300	334	340	265	265	207	134	48	.	.	.

C) PROBLEMS

1. Convert panels A and B to coded form. Compare.

2. As problem 2 of exhibit 21.

exhibit 23 of chapter 29x

Table of double flogs and log smallest corners for
the biGaussian pattern of distribution shown in exhibit
strong relationship -- $\rho = 0.9$

A) The DOUBLE FLOGS - raw

											(flog)
.	.	.	.	.	.	.	3.94	2.97	2.54	2.55	(2.20)
.	.	.	.	.	.	3.84	2.85	2.22	2.17	2.54	(1.63)
.	.	.	.	.	3.72	2.70	1.99	1.89	2.22	2.97	(1.15)
.	.	.	.	3.56	2.51	1.85	1.70	1.99	2.85	3.94	(.73)
.	.	.	3.56	2.49	1.79	1.58	1.85	2.70	3.84	.	(.35)
.	.	3.72	2.51	1.79	**1.55**	1.79	2.51	3.72	.	.	(0)
.	3.84	2.70	1.85	1.58	1.79	2.49	3.56	.	.	.	(-.35)
3.94	2.85	1.99	1.70	1.85	2.51	3.56	.	.	.	.	(-.73)
2.97	2.22	1.89	1.99	2.70	3.72	.	.	.	.	.	(-1.15)
2.54	2.17	2.22	2.85	3.84	.	.	.	.	.	.	(-1.63)
2.55	2.54	2.97	3.94	.	.	.	.	.	.	.	(-2.20)
(flog) (-2.20)	(-1.63)	(-1.15)	(-.73)	(-.35)	(0)	(.35)	(.73)	(1.15)	(1.63)	(2.20)	

B) The LOG SMALLEST CORNERS -- unit = 0.01

.	.	.	.	.	.	.	.60	2.00	2.94	3.47
.	.	.	.	.	.	.85	2.39	3.37	3.97	2.94
.	.	.	.	.	1.15	2.60	3.73	4.36	3.37	2.00
.	.	.	.	1.30	2.85	3.95	4.64	3.73	2.39	.60
.	.	.	1.30	2.91	4.07	4.80	3.95	2.60	.85	.
.	.	1.15	2.85	4.07	**4.86**	4.07	2.85	1.15	.	.
.	.85	2.60	3.95	4.80	4.07	2.91	1.30	.	.	.
.60	2.39	3.73	4.64	3.95	2.85	1.30	.	.	.	.
2.00	3.37	4.36	3.73	2.65	1.15	.	.	.	.	.
2.94	3.97	3.37	2.39	.85	.	.	.	.	.	.
3.47	2.94	2.00	.60	.	.	.	.	.	.	.

C) PROBLEMS

1. As problem 2 of exhibit 21.

2. Why do we omit the analog of problem 1 of exhibits 21 or 22?

exhibit 24 of chapter 29x

Table of double flogs and log smallest corners for
the biGaussian pattern of distribution shown in exhibit
Very strong relationship -- $\rho = 0.95$

A) The DOUBLE FLOGS -- raw

											(flog)
.	.	.	.	.	.	.	.	4.49	3.41	2.92	(2.20)
.	.	.	.	.	.	.	4.24	2.85	2.53	3.41	(1.63)
.	.	.	.	.	.	4.12	2.62	2.24	2.85	4.49	(1.15)
.	.	.	.	.	3.97	2.48	2.04	2.62	4.24	.	(.73)
.	.	.	.	3.97	2.41	1.93	2.48	4.12	.	.	(.35)
.	.	.	3.87	2.41	**1.90**	2.41	3.97	.	.	.	(0)
.	.	4.12	2.48	1.93	2.41	3.97	.	.	.	.	(-.35)
.	4.24	2.62	2.04	2.48	3.97	.	.	.	.	.	(-.73)
4.49	2.85	2.24	2.62	4.12	.	.	.	.	.	.	(-1.15)
3.41	2.53	2.85	4.24	.	.	.	.	.	.	.	(-1.63)
2.92	3.41	4.49	.	.	.	.	.	.	.	.	(-2.20)
(flog) (-2.20)	(-1.63)	(-1.15)	(-.73)	(-.35)	(0)	(.35)	(.73)	(1.15)	(1.63)	(2.20)	

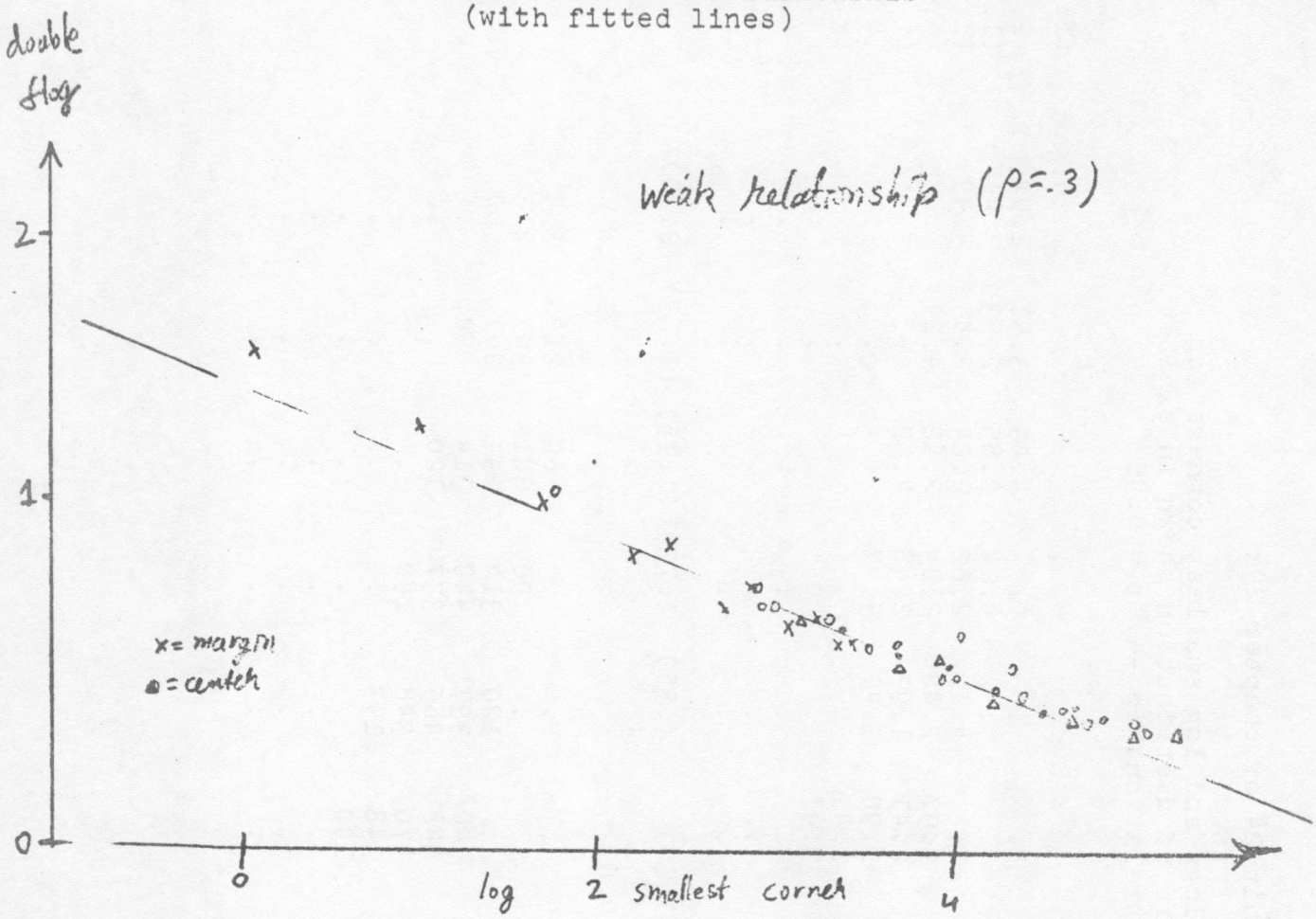
B) The LOG SMALLEST CORNERS -- unit = 0.01

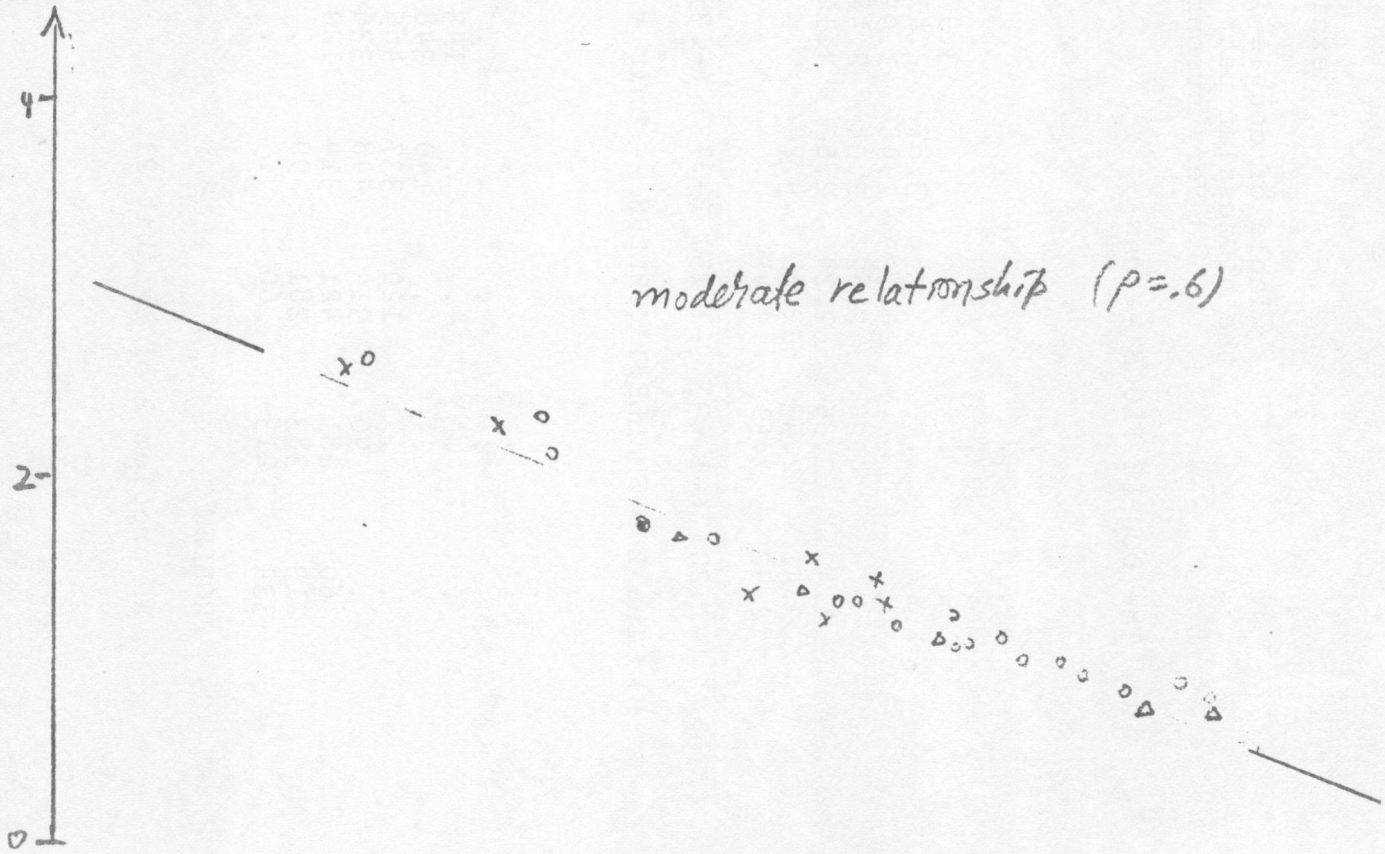
.	.	.	.	.	.	.	.	60	213	334
.	.	.	.	.	.	.	90	281	303	213
.	.	.	.	.	.	120	314	241	281	60
.	.	.	.	.	140	337	448	314	90	.
.	.	.	.	143	348	465	337	120	.	.
.	.	.	140	348	**470**	348	140	.	.	.
.	.	120	337	465	348	193	.	.	.	.
.	90	314	448	336	140	.	.	.	.	.
60	281	421	314	120	.	.	.	.	.	.
213	383	281	90	.	.	.	.	.	.	.
334	213	60	.	.	.	.	.	.	.	.

C) PROBLEM

As problem 2 of exhibit 21.

Plots of double flog against log smallest corenr
for weak and moderate relationship
(with fitted lines)





Plots of double flogs against log smallest corner
for strong and very strong relationship
(with fitted lines)

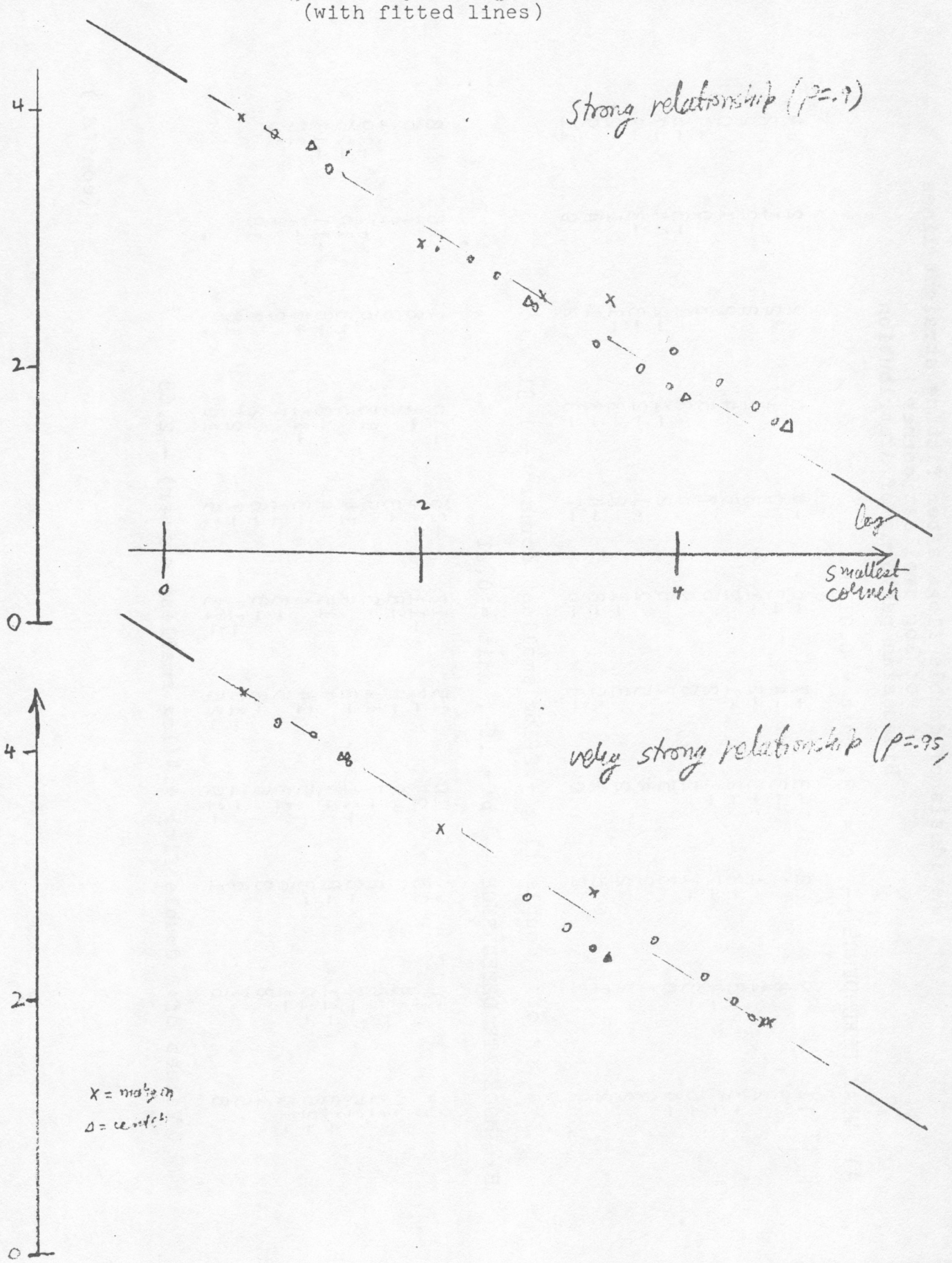


exhibit 27 of chapter 29x

Residuals of double flogs after fitting straight lines
on log smallest corner
BiGaussian patterns of distribution

A) WEAK DEPENDENCE -- ρ = .3, unit = 0.01

11	9	2	-3	-7	-6	-4	0	2	2	4
9	4	1	-3	-4	-3	0	1	2	1	2
2	1	-1	-3	-2	-1	3	2	2	2	2
-3	-3	-3	-2	-1	2	5	4	2	1	0
-7	-4	-2	-1	2	6	7	5	3	0	-4
-6	-3	-1	2	6	9	6	2	-1	-3	-6
-4	0	3	5	7	6	2	-1	-2	-4	-7
0	1	2	4	5	2	-1	-2	-3	-3	-3
2	2	2	2	3	-1	-2	-3	-1	1	2
2	1	3	1	0	-3	-4	-3	1	4	9
4	2	2	0	-4	-6	-7	-3	2	9	11

Values of: double flog +.2(log smallest corner) - 1.31

B) MODERATE DEPENDENCE -- ρ = .6 , unit = 0.01

.	.	.	10	15	-13	-26	-19	1	6	8
.	.	18	24	-7	-11	-10	-1	8	7	6
.	18	7	2	-8	-8	-3	6	8	8	1
10	24	2	-7	-11	-3	5	13	6	-1	-19
15	-7	-8	-11	-3	4	14	5	-3	-10	-26
-13	-11	-8	-3	4	15	4	-3	-8	-11	-13
-26	-10	-3	5	14	4	-3	-11	-8	-7	15
-19	-1	6	13	5	-3	-11	-7	2	24	10
1	8	8	6	-3	-8	-8	2	7	18	.
6	7	8	-1	-10	-11	-7	24	18	.	.
8	6	1	-19	-26	-13	15	10	.	.	.

Values of: double flog +.4(log smallest corner) - 2.68

(cont'd.)

exhibit 27 (cont'd.)

C) STRONG DEPENDENCE -- $\rho = 0.9$, unit = 0.01

.	.	.	.	.	.	.	-1	-14	-1	32
.	.	.	.	.	.	4	-3	-7	24	-1
.	.	.	.	.	10	-5	-8	20	-7	-14
.	.	.	.	3	-8	-8	20	-8	-3	-1
.	.	.	3	-7	-8	15	-8	-5	4	.
.	.	10	-8	-8	16	-8	-8	10	.	.
.	4	-5	-8	15	-8	-7	3	.	.	.
-1	-3	-8	20	-8	-8	3	.	.	.	.
-14	-7	20	-8	5	10	.	.	.	.	.
-1	24	-7	-3	4	.	.	.	.	.	.
32	-1	-14	-1	.	.	.	.	.	.	.

Values of: double flog +.6(log smallest corner) - 4.31

D) VERY STRONG DEPENDENCE -- $\rho = .95$, unit = 0.01

.	.	.	.	.	.	.	.	2	-10	17
.	.	.	.	.	.	.	-4	-23	10	-10
.	.	.	.	.	.	3	-25	4	-23	2
.	.	.	.	.	0	-24	2	-25	-4	.
.	.	.	.	2	-25	2	-24	3	.	.
.	.	.	0	-25	2	-25	0	.	.	.
.	.	3	-24	2	-25	2	.	.	.	.
.	-4	-25	2	-24	0	.	.	.	.	.
2	-23	4	-25	3	.	.	.	.	.	.
-10	10	-23	-4	.	.	.	.	.	.	.
17	-10	2	.	.	.	.	.	.	.	.

Values of: double flog +.63(log smallest corner) - 4.85

(cont'd.)

29x-62

exhibit 27 (cont'd.)

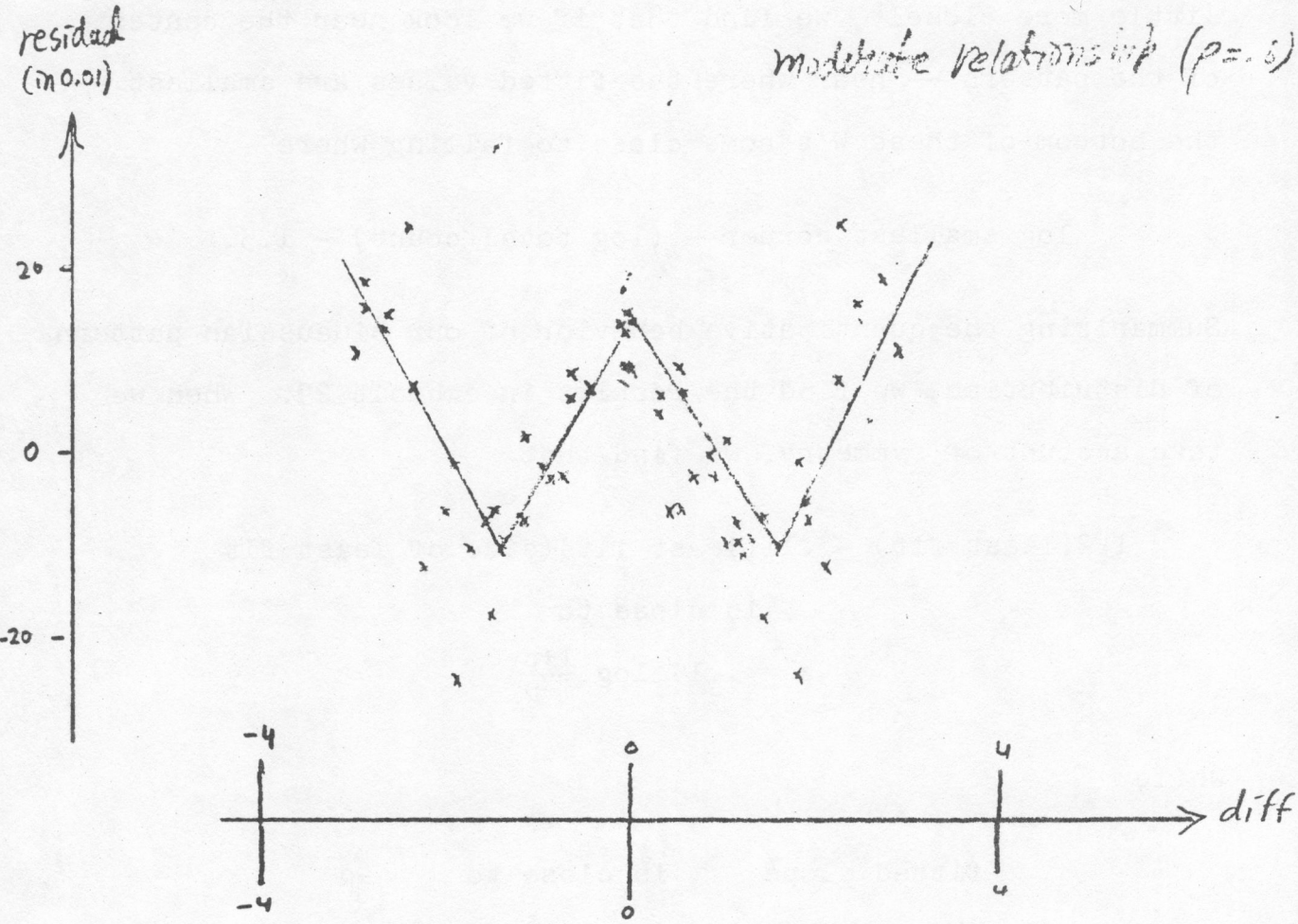
E) VALUES of COL FLOG minus ROW FLOG

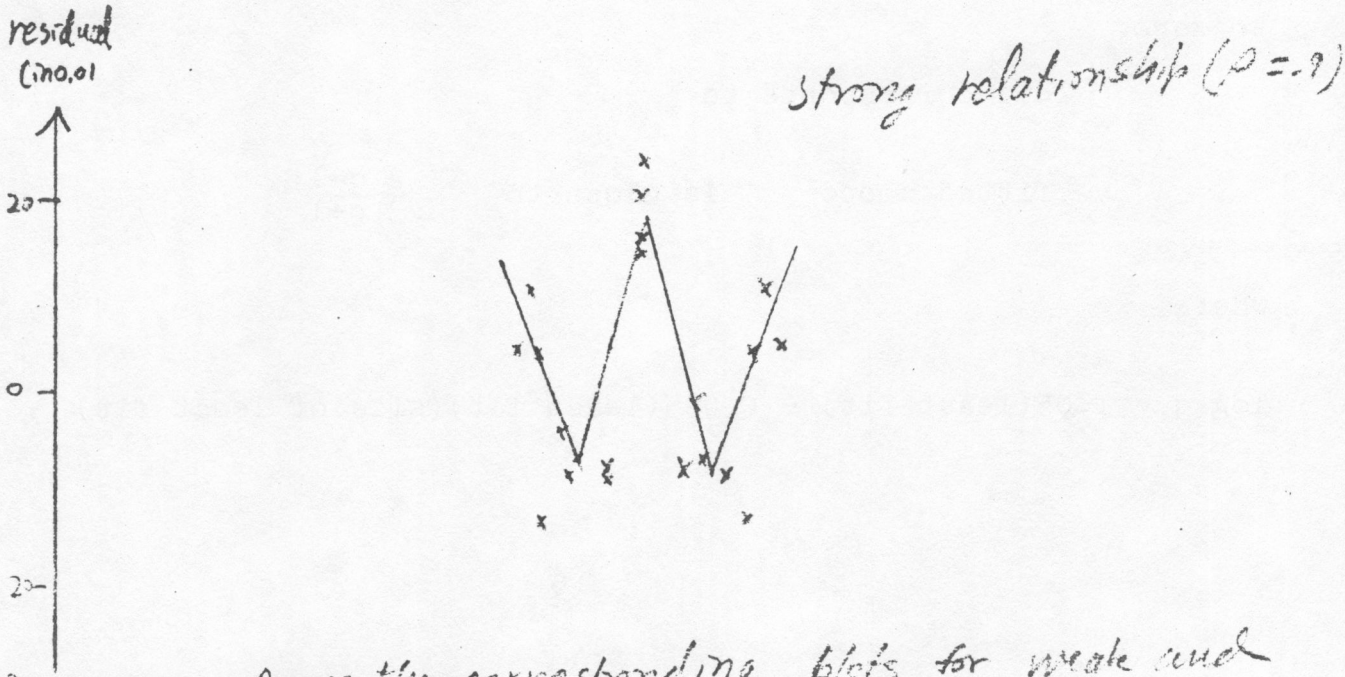
-440	-383	-335	-293	-255	-220	-185	-147	-105	-57	0
-383	-326	-278	-236	-198	-163	-128	-90	-58	0	57
-335	-278	-230	-188	-150	-115	-80	-42	0	58	105
-293	-236	-188	-146	-108	-73	-38	0	42	90	147
-255	-198	-150	-108	-70	-35	0	38	80	128	185
-220	-163	-115	-73	-35	0	35	73	115	163	220
-185	-128	-80	-38	0	35	70	108	150	198	255
-147	-90	-42	0	38	73	108	146	188	236	293
-105	-58	0	42	80	115	150	188	230	278	335
-57	0	58	90	128	163	198	236	278	326	383
0	57	105	147	185	220	255	293	335	383	440

F) PROBLEM

Replace panel E, and at least two of panels A to D by coded values. (Even guesses at hinges will suffice.)

Plots of residual against "diff" for moderate and strong relationship with eye-fitted W's





PROBLEM: Draw the corresponding plots for weak and very strong relationships

When we check into this behavior of residuals a little more closely, we find that if we look near the center of the pattern -- near where the fitted values are smallest -- the bottom of these W's come close to falling where

log smallest corner - (log total count) - 1.5.

Summarizing the quantitative behavior of our biGaussian patterns of distribution, we find the results in exhibit 29. When we take account of symmetry, we find that

1.2(least fit) - .11(least fit)(size of least fit

is close to

$$z = 1.15 \log \frac{1+\rho}{1-\rho}$$

while

fitted slope **is close to** $\frac{2}{3}\rho$

where "least fit" refers to the fitted value of ff-log closest to zero.

This corresponds to

fitted slope **is close to** $\frac{2}{3}\,\frac{q-1}{q+1}$

where

log q = 1.05(least fit) - 0.095(least fit)(size of least fit)

exhibit 29 of chapter 29x

Quantitative relation of biGaussian patterns
of distribution to the fit on log smallest corner

Strength of relationship:	None	Weak	Moderate	Strong	Very strong
Lowest* ff-log:	0	.34	.76	1.55	1.90
Least* fitted value:	0	.25	.61	1.31.39	1.89
1.2(least fit)-.11(least fit)^2	0	.30	.69	1.46	1.85
Fisher's z:	0	.31	.69h	1.47	1.83
ρ	0	.3	.6	.9	.95
$\frac{2}{3}\rho$	0	.2	.4	.6	.63
fitted slope	0	.2	.4	.6	.63
W for residuals:					
center height:	0	.08	.14	.18	.08
depth:**	0	-.04 (at ±2.0)	-.10 (at ±1.5)	-.09 (at ±.7)	-.24 (at ±0.4)
end height**:	0	.09 (at ±4.0)	.20 (at ±3.0)	.07 (at ±1.4)	.07 (at ±0.8)

* Occurs at (or near) col flog = 0 = row flog

**At the values of diff = col flog - row flog indicated in parentheses.

a relation we can use to see how biGaussian -- in one aspect -- an actual pattern of distribution seems to be.

For our bean dimensions, as can be calculated from exhibits 14 and 15, we have:

fitted slope = 4.3 ; least fit = 11.7

for which

log q = 1.23 - .13 = 1.10 q = 12.6

so that the relation becomes

.43 may be close to .57

This suggests that the behavior of the bean dimensions is close to biGaussian, but probably somewhat different.

29xF. <u>The bean dimensions further fitted</u>

When we came to the clearly structured residuals of exhibit 15, we postponed further analysis to "a later section" in order to bring in the material of the last section, which is more closely related to the main purpose of this chapter. We still feel obliged, however, to carry the analysis of the bean dimension data to a reasonable stopping point, which we shall do in the present section.

When we come to still more detailed analyses of a two-way table of double flogs, there are four carriers -- three

of which we have already met -- that are likely to be particularly useful (in addition, perhaps to col flog and row flog themselves). We shall feel free to use the following abbreviations:

smal = log smallest corner

diff = col flog MINUS row flog

sum = col flog PLUS row flog

prod = (col flog)(row flog)

We are by now familiar with the patterns of values for each of these carriers but "sum", whose pattern is that of a gradation from both flogs low to both flogs high

* the 2nd step *

When we turn back to exhibit 15 we see the same general sort of pattern among the residuals that we expect for "sum" -- low at one end, high at the other. Indeed, if we define our column and row flogs as in panel C of exhibit 14, that + flogs go with small dimensions (some might rather use opposite signs), we find the direction of residuals and "sum" the same. It is natural to make the removal of a tilt on sum our next step.

Exhibit 30 shows the values of

sum = col flog PLUS row flog

and the 2nd residuals found from

2nd residuals = residuals MINUS .1(sum).

exhibit 30 of chapter 29x -- bean dimensions

The values of "sum" and of the 2nd residuals of
bean-dimension double flogs after taking out a tilt on sum

A) Values of COL FLOG plus ROW FLOG, divided into NINTHS -- unit = 0.01

	-596	-465	-398									
-545	-443	-362	-295									
	-348	-267	-260	-152	-109							
	-277	-196	-129	-81	-38	3						
		-138	-71	-23	20	61	98	129				
		-85	-18	30	73	114	151	182	209			
				82	126	166	203	234	261	289		
					176	217	254	285	312	340	373	44	462
							303	334	361	389	422	460	511
									440	468	501	539	590
									494	522	555	593	644

Median: 151. Hinges: -85 and 373. 8ths: -286 and 497.

B) The NINE MEDIAN SUMMARY POINTS -- based on panel C of exhibit 15.

(-443,-25), (-198,-26), (-85,-22), (61,-10), (151,2), (254,24), (340,40), (461,38), (555,69)

giving end thirds of (-242,-24) and (452,49) suggesting a tilt of 0.1 = 73/694.

(cont'd.)

exhibit 30 (cont'd.)

C) Values of 2ND RESIDUAL -- residual MINUS .1(panel A), unit still 0.01

a	30	-55	-132									
25	37	14	8									
	12	24	-2	-14	1							
	-10	-10	1	-7	-18	-16						
		-22	-16	-7	-8	-18	-15	-11				
		-60	-49	-22	-14	-10	-17	-14	8			
				-18	-18	-16	2	1	-2	19		
					6	-19	-5	8	-19	8	-5	19	-46
							26	-3	4	-4	9	2	-19
									-13	-21	-6	49	45
									-83	-58	-45	10	33

Median: -6. Hinges: -18 and 8. 8ths: -34± and 22. Outs: 11/66. Det: 2/66.

D) Coded values for 2ND RESIDUALS

.	x	0	0									
x	X	x	.									
	x	x	.	.	.							
	.	.	.	.	.	.						
		o	.	.	.	.	.					
		0	0	o	.	.	.	.	.			
				.	.	.	.	.	.	x		
					.	o	.	.	x	.	.	x	0
							x	.	.	.	.	.	o
									.	o	.	X	X
									0	0	0	x	X

Median: -6. Hinges: -18 and 8. 8ths: -34± and 22. Outs: 11/66. Det: 2/66.

E) PROBLEM

How much difference would it have made to keep one less figure in panel A and the subsequent fitting? Why?

29x-70

This further fit reduces the midspread from .54 to .26, and is clearly a major improvement.

We thus appear to have a stronger relationship among bean dimensions when they are smallish than when they are largish. The cause for this is an interesting botanical query. Do such things happen routinely? Do the 9440 beans contain sufficiently many pathologically undersized ones whose removal would eliminate any tendency for double flog to trend with sum? Or what?

* the 3rd step *

When we look at the 2nd residuals in coded form (panel D), we see a slight tendency toward the sort of behavior that we saw for biGaussian patterns of distribution in the last section -- diagonal streakiness with values on the diagonal higher than values near it. It is natural then, to try to fit such a "W" pattern.

The values in panel A of exhibit 31 were found by interpolating in exhibit 29 to a least fitted value of 1.14 - (-.03) = 1.17 which falls almost exactly midway between the columns for "moderate" and "strong" relationship. Simple interpolation gives +.16 at 0, -.10 at ±1.1 and +.14 at ±2.2. In exhibit 31 all these have been decreased by .16, so as to start at (0,0). The 9 summary points are given in panel B of exhibit 31 and plotted in exhibit 32.

exhibit 31 of chapter 29x -- bean dimensions

The W-shaped fit and the third residuals

A) The W-SHAPED FIT -- unit 0.01

a	-26	-8	6									
-23	-2	-21	-16									
	-21	-2	-14	-26	-17							
	-15	-19	-3	-9	-19	-24						
		-20	-16	-5	-5	-15	-24	-21				
		-9	-27	-18	-7	-2	-11	-19	-25			
				-22	-20	-10	-1	-6	-13	-19		
					-21	-22	-13	-6	0	-7	-15	-24	-17
							-25	-18	-11	-4	-3	-13	-25
									-22	-23	-15	-6	-6
									-11	-17	-25	-19	-7

Values found from:

-.238(size of diff) for $-1.1 \leq \text{diff} \leq 1.1$

-.50 + .218(size of diff) elsewhere

B) The 9 SUMMARY POINTS -- (panel A above, panel C of exhibit 30)

(-26,8), (-23,-16), (-21,-2), (-19,-10), (-16,-10), (-13,-2), (-8,-14), (-5,1), (-2,10)

(cont'd.)

exhibit 31 (cont'd.)

C) The 3RD RESIDUALS -- unit still 0.01

a	56	-47	-138										
48	39	35	24										
	33	26	12	12	18								
	5	9	4	2	1	8							
		-2	0	-2	-3	-3	9	10					
		-51	-28	-4	-7	-8	-6	5	33				
				4	2	-6	3	7	15	38			
					27	3	8	14	19	15	10	43	-29
							51	15	15	0	12	15	6
									9	2	9	55	52
									-72	-41	-20	29	40

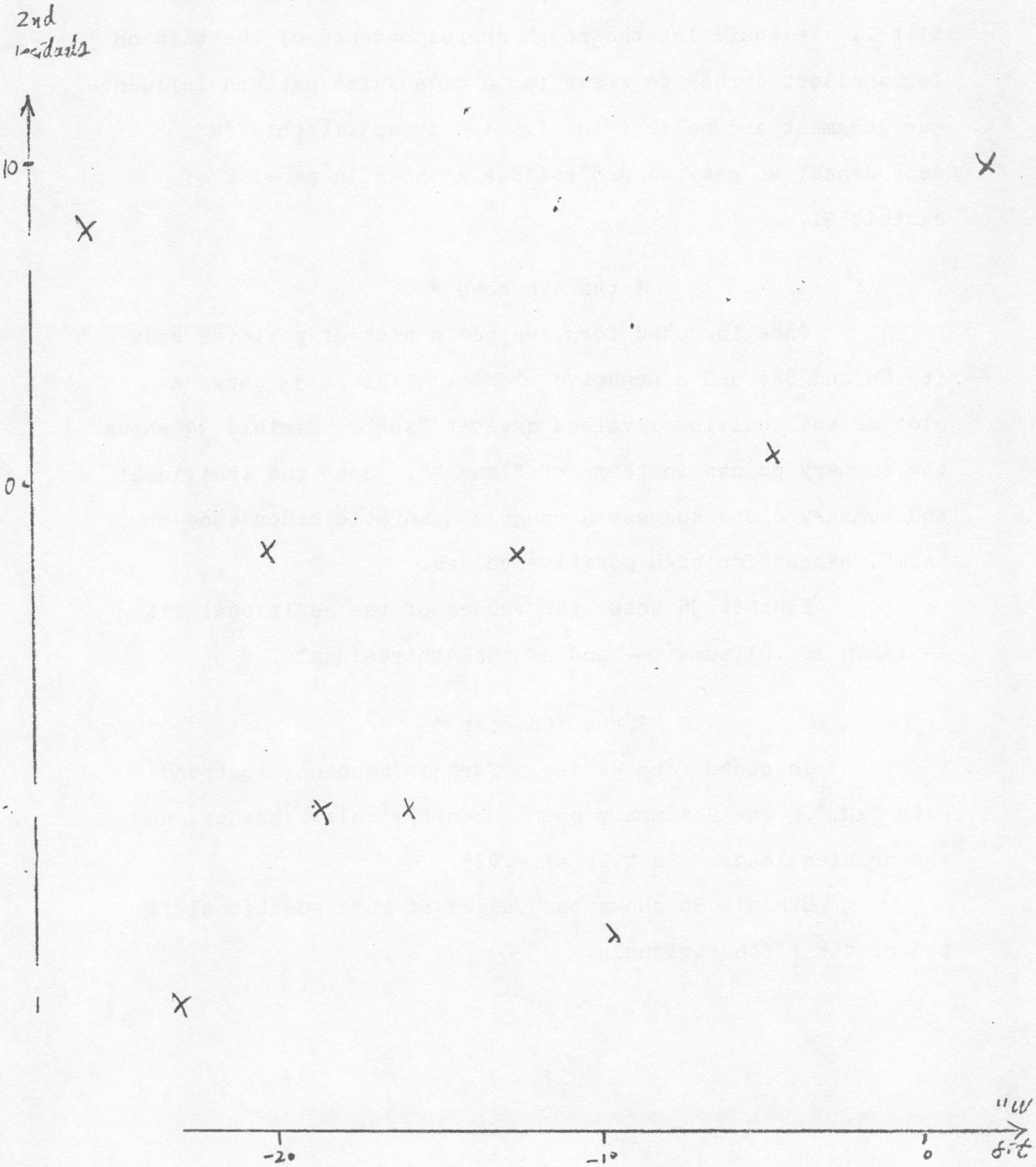
Median: 8. Hinges: -2 and 24. 8ths: -13 and 38. Outs: 10/66. Dets: 2/66

D) The 3RD RESIDUALS CODED

a	0	0	0										
X	X	X	x										
	X	x	.	.	x								
	.	.	.	.	.	.							
		.	.	.	.	.	.	.					
		0	0	.	o	o	o	.	.				
				.	.	.	.	0	X	.			
					x	X	.	.	o	.	.	0	0
							X	.	x	.	.	.	
									o	.	x	X	X
									0	0	0	.	X

29x-73

Plot of the nine summary points for 2nd residuals
against the W-shaped "fit"
(see panel B of exhibit 31)



It is almost equally easy to believe in tilt 0 or tilt 1. We shall let the rough correspondence of the tilt on log smallest corner to treat for a biGaussian pattern influence our judgment and believe in tilt 1. If we fit this "W" dependence, we get the 3rd residuals shown in panel C of exhibit 31.

* the 4th step *

Once in coded form, we see a hint of positive ends (to NW and SE) and a negative middle. Exhibit 33 shows a plot of the individual values against "sum". Exhibit 34 shows the summary points in terms of "$(\text{sum})^2$". Both the individual and summary plots suggest a roughly quadratic dependence on "sum", except for high positive values.

Exhibit 35 shows the values of the additional fit -- taken as $.01(\text{sum})^2$ -- and of the 4th residuals.

* the 5th step *

In coded form we see a further tendency to trend with "sum". The 9 summary points confirm this tendency, and the problem leads to a tilt of -.025.

Exhibit 36 shows the values of this additional fit and of the fifth residuals.

Plot of 3rd residuals against "sum", individual values and summary points

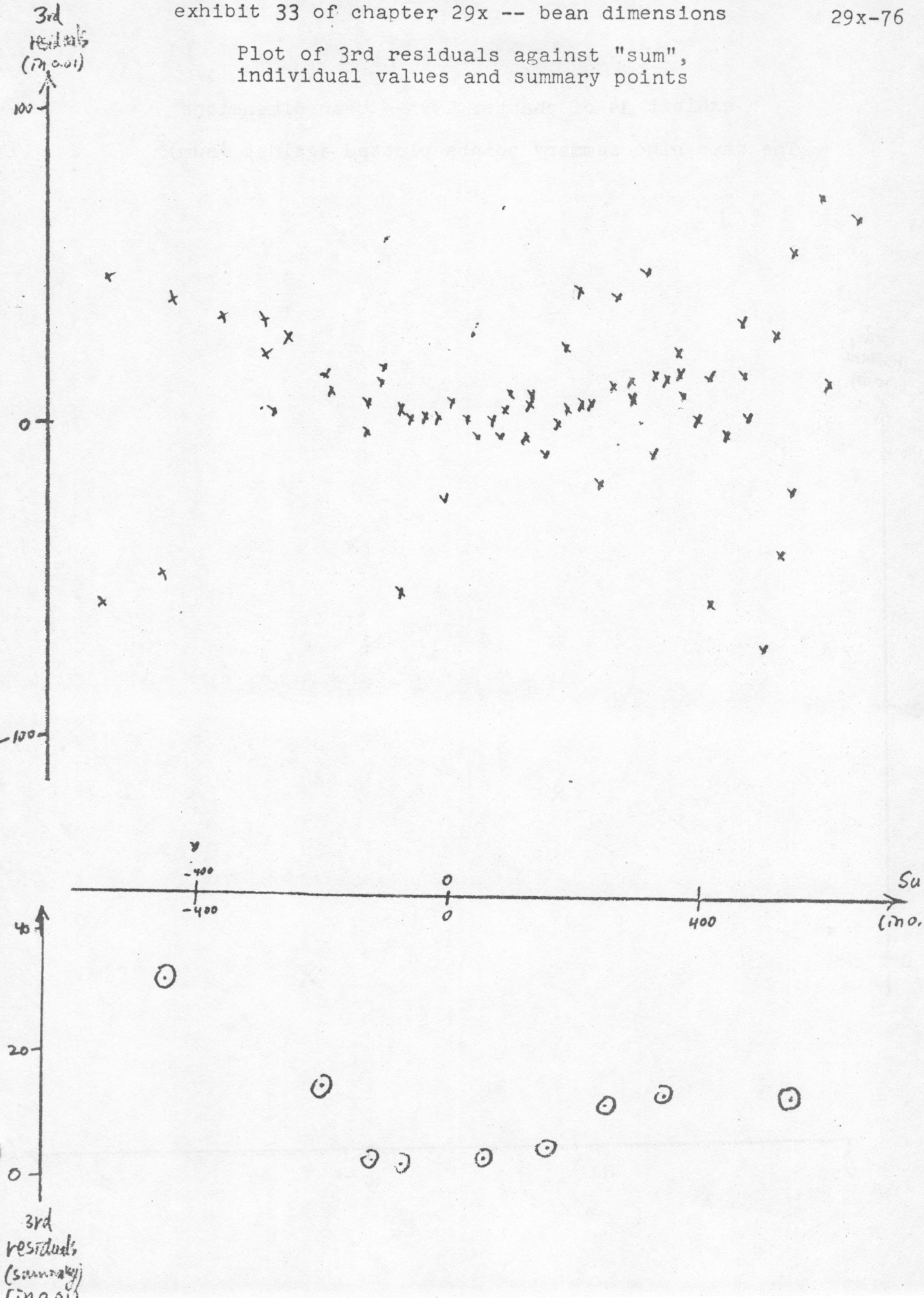


exhibit 34 of chapter 29x -- bean dimensions

The same nine summary points plotted against $(\text{sum})^2$

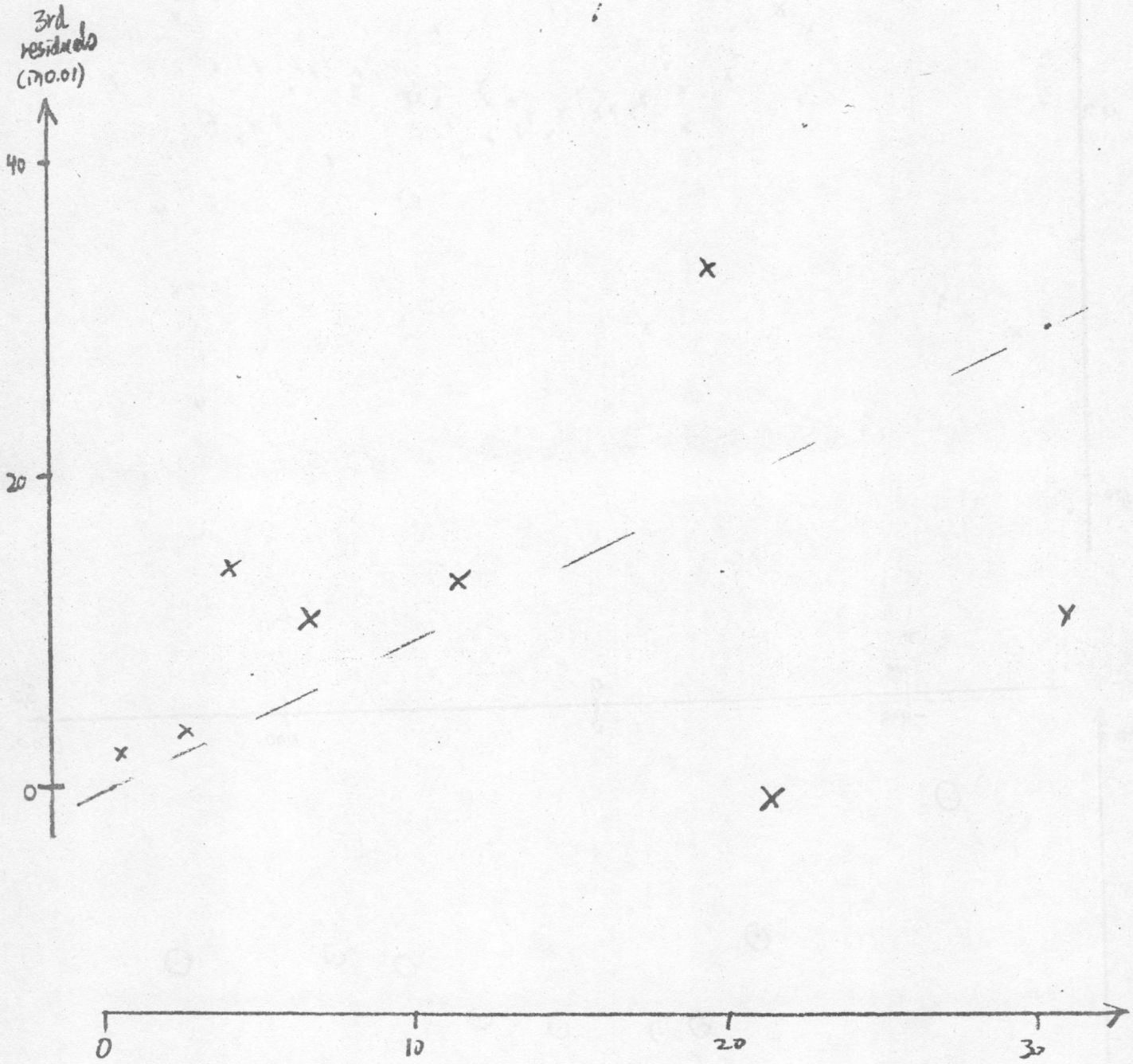


exhibit 35 of chapter 29x -- bean dimensions

Removal of a fit on $(\text{sum})^2$

A) Values of $.01(\text{SUM})^2$ -- the additional fit

	30	22	16									
30	20	13	9									
	12	7	4	2	1							
	8	4	2	1	0	0						
		2	1	0	0	0	1	2				
		1	0	0	1	1	2	3	4			
				1	2	3	4	5	7	8		
					3	5	6	8	10	12	14	17	21
							9	11	13	15	18	21	26
									19	22	25	29	35
									24	27	31	35	41

B) Values of 4TH RESIDUALS = 3RD RESIDUALS minus ABOVE

	26	-69	-154										
16	19	23	15										
	21	19	8	10	-19								
	-3	5	2	1	1	0							
		-4	-1	-2	-3	-3	8	8					
		-52	26	-4	-8	-9	-8	2	29				
				3	0	-9	-1	2	8	30			
					24	-2	2	6	9	3	-4	-26	-50
							42	4	2	-15	-6	-6	-20
									-10	-20	-16	26	17
									-96	-68	-51	-6	-1

Median: 4. Hinges: -8 and 8. 8ths: -23 and 22. 16ths: -40 and 26. Outs: 14/66.

(cont'd.)

exhibit 35 (cont'd.)

C) CODED values of 4TH RESIDUALS

```
        X     ?     ?
x       x     x     x
        x     x     .     x     o
        .     .     .     .     .     .
              .     .     .     .     .     .
              ?     X     .     .     o     .     .     X
                          .     .     o     .     .     .     X
                                x     .     .     .     x     .     .     0     ?
                                            X     .     .     o     .     .     o
                                                        o     o     o     X     x
                                                        ?     ?     ?     .     .
```

Note: Values at or beyond ±50 coded as ?

D) SUMMARY POINTS, according to "SUM", for 4th RESIDUALS -- unit still 0.01

sum	4th residuals	summary points
-546 to -348	16, 26, 19, 21, 23, -69, -154	(-443,19)
-295 to -129	-3, -4, 5, 19, 2, 8, 15, 10	(-198,8)
-109 to -18	26, -1, -2, 1, 1, -19, -52	(-85,-1)
3 to 98	3, -4, -8, -3, -3, 0, 8	(61,-3)
114 to 203	24, 0, -9, -1, -9, 8, 2, 8	(151,1)
209 to 289	-2, 2, 6, 2, 8, 29, 30	(254,6)
303 to 389	4, 2, -15, 9, 3, -4, 42	(
410 to 501	-10, -20, -16, -6, -6, -26, -96, -50	(461,-18)
511 to 644	-6, -1, 26, 17, -20, -68, -51	(555,-6)

E) PROBLEM

Plot the summary points of panel D and confirm that a tilt of -.025 is indicated.

29x-79

exhibit 36 of chapter 29x -- bean dimensions

Another additional fit and the 5th residuals

A) The ADDITIONAL FIT -- given by -.025(sum)

a	14	12	10									
14	11	9	7									
	9	7	5	4	3							
	7	5	3	2	1	0						
		4	2	1	0	-2	-2	-3				
		2	0	-1	-2	-3	-4	-5	-5			
				-2	-3	-4	-5	-6	-7	-7		
					-4	-5	-6	-7	-8	-8	-9	-10	-12
							-8	-8	-9	-10	-11	-12	-13
									-11	-12	-13	-13	-15
									-12	-13	-14	-15	-16

B) Values of 5TH RESIDUAL = 4TH RESIDUAL minus panel A

	12	-81	-164										
2	8	14	8										
	12	12	3	6	-22								
	-10	0	-1	-1	0	0							
		-7	-3	-3	-3	-1	10	11					
		-54	26	-3	6	-6	-4	7	34				
				5	3	-5	4	8	15	37			
					28	3	8	13	17	11	5	-16	-38
							50	12	11	-5	5	6	-7
									1	-8	-4	39	32
									-84	-55	-37	9	15

Median: 4. Hinges: -4 and 11. 8ths: -19 and 16. 16ths: -54 and 31. Outs: 14/66.

(cont'd.)

exhibit 36 (cont'd.)

C) CODED values of 5th RESIDUALS

```
        x     ?     ?
.       .     x     .
        x     x     .     .     0
        o     .     .     .     .
              o     .     .     .     .     .
              ?     x     .     .     o     .     .     X
                          .     .     o     .     .     x     X
                                X     .     .     x     x     .     .     o     0
                                            ?     x     .     .     .     .     o
                                                        .     o     .     X     X
                                                        ?     ?     0     .     x
```

Note: Values at and beyond ±50 coded ?

29x-81

* the 6th step *

The pattern of the 5th residuals is not too clear. It seems almost time to take stock of where we have been and quit. Before doing this, however, it seems sensible to ask about the first steps of fitting we took whether we can now do better. Our first step was to fit log smallest corner. Exhibits 37 and 38 show what happens when we look at the relation of the 5th residuals to this carrier.

The crucial display is exhibit 38, which shows both summary points and somewhat schematized information about the individual values. In the right hand three-fifths of this picture, where the smallest corner contains 10 or more beans, the residuals are closely packed together and suggest a slight negative slope.

We have previously acted as if some values were rather straying. If the low values for smallest corners of 1 and 2 are thought of as strays, a gentle negative tilt would be quite compatible with exhibit 38. Panels B and C of exhibit 37 deal with the corresponding additional fit.

* strays? *

Where do these unusually negative residuals come from? Panel A of exhibit 39 shows the portions of the original tables of counts from which the five most negative residuals come. In every instance there is a cell with a nonzero count which appears

exhibit 37 of chapter 29x -- bean dimensions

Back to log smallest corner again; 6th residuals

A) 9 SUMMARY POINTS -- for (log smallest corner, 5th residuals), units still 0.01

log smallest corner	5th residuals	Summary points
0	-54, 28, 56, 1, -84, -55, 9, -16, 9, -38	(0, -8)
30	2, 12, -81, -54, -164, -22, -7, 15	(30,-14)
48	8, 11, 34, -8, 39, 32	(48, 22)
60 to 90	-10, -7, 26, 5, 37, 6, 39	(70, 6)
95 to 118	8, 0, 3, 10, 12, 5	(112, 6)
128 to 152	12, 14, 6, 3, 8, 7, 11, 5	(144, 8)
153 to 188	0, -3, 0, 13, 15, 11, -5	(175, 0)
197 to 220	12, 3, -3, -1, -5, -4, 8, 17	(206, 0)
246 to 274	-1, -1, -3, 6, -6, 4	(271, -1)

Turn to exhibit 38 before going on to panel B.

B) Values of .10 - .04 log smallest corner

	9	9	9									
9	5	4	6									
	5	1	2	4	9							
	8	3	-1	0	3	6						
		6	2	-1	-1	2	5	8				
		10	7	2	0	-1	2	4	8			
				7	4	1	0	2	4	8		
					10	5	4	3	2	4	6	10	10
							10	6	5	4	4	6	9
									10	8	8	8	8
									10	10	10	10	9

(cont'd)

exhibit 37 (cont'd.)

C) Values of 6TH RESIDUALS == 5TH RESIDUALS minus ABOVE

	3	-90	-173									
-7	3	10	2									
	7	11	1	2	-31							
	-18	-3	0	-1	-3	-6						
		-13	-5	-2	-2	-3	5	3				
		-64	19	-5	6	-5	-6	3	26			
				-2	-1	-6	4	6	11	29		
					18	2	4	8	15	7	-1	-26	-48
							46	2	6	-10	0	0	-16
									-9	-16	-12	31	24
									-94	-65	-47	-1	6

Median: -1. Hinges: -7 and 6. 8ths: -28 and 13. 16ths: -64 and 25. Outs: 14/66.

Note: If we set aside 5 entries of -64 and beyond, the summary for the remaining 61 becomes:

Median: 0. Hinges: -6 and 6. 8ths: -13 and 15. 16ths: -26 and 26.

29x-84

Plot of 5th residuals against log smallest corner
Individual residuals collected at summary values of carrier
Dashed horizontals pass through summary points

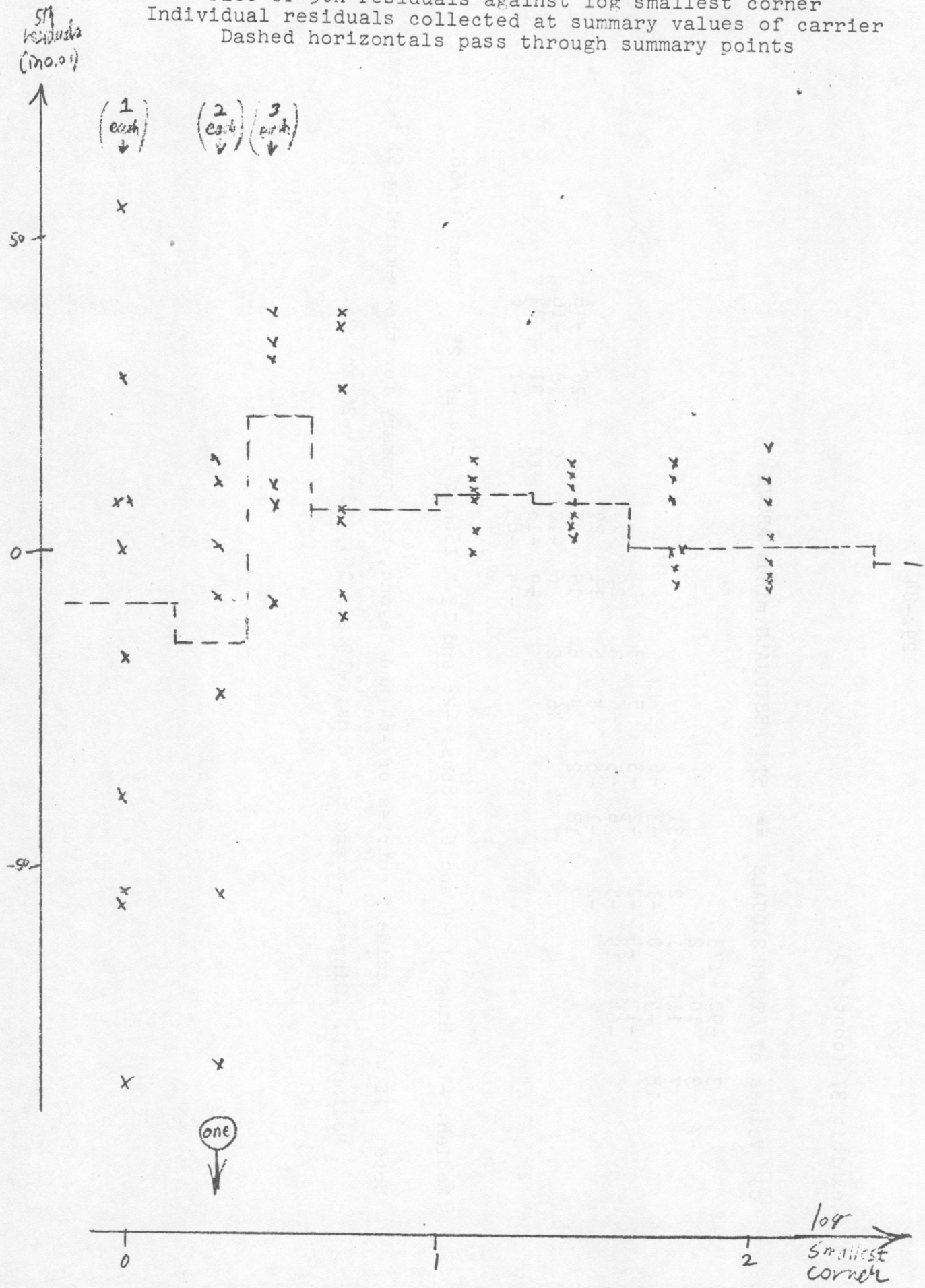


exhibit 39 of chapter 29x -- bean dimensions

The sources of the strays and the coded 6th residuals

A) WHERE do the 6th residuals BEYOND -60 COME FROM? -- parts of original table of counts from exhibit 14.

-90 and -173

.	2	.	.	3		(smallest corner = 2)
4	8	17	19	.		
2	23	101	156	93	23	2

-64

4	44	375	956
	7	81	385
	1	4	65
			6

-94 and -65

1	13	18	28	35	25	32
		1	9	8	21	12
				.	2	.
				1	.	.

B) Coded values of 61 other 6TH RESIDUALS -- the set aside 5 coded as "?"

.	.	?	?									
o	.	x	.									
	x	x	.	.	0							
	0	.	.	.	.	.						
		o	.	.	.	.	.	.				
		?	x	.	.	.	.	.	X			
				.	.	.	.	.	x	X		
					x	.	.	x	x	.	.	0	0
							X	.	.	o	.	.	o
									o	o	o	X	X
									?	?	0	.	.

C) PROBLEM

Go back to exhibit 31, do NOT take out the W-shaped fit, but follow through the rest of the path just taken.

quite isolated. In the case of the first and 3rd cases, near the NW and SW ends of the pattern, respectively, we can be relatively sure that these values are strays -- meaning by this that if they would recur in other batches of comparable size, then either:

- they would involve demonstrably unusual measurement errors, or
- they would be likely to be distinguishable on qualitative biological grounds.

The second case, yielding -64, would not seem to be so clearly a stray, but the case with which one bean produces a large residual warns us not to take the -64 too seriously.

The analysis of panel B probably represents as well as we can do by our path.

* comparison of steps *

Exhibit 40 summarizes our fitting path. Clearly the gains through step 4 are major. The gain from step 4 to step 5 appears definite. Beyond that, differing relative attention to 8thspread and midspread can lead to differing choices. The writer leans toward "(***)" with 5 double flogs treated as strays. Either of the other two choices "after 5" and "after 6" can, of course, be argued for.

Whichever of these three we choose as a stopping point, we must judge our appearances as saying:

exhibit 40 of chapter 29x -- bean dimensions

Our path of fitting to the double flogs of the bean-dimension contingency table summarized.
(66 values to solve)

A) The STEPS

#	fitted	coefficients added: 1	small	sum	"W"	$(sum)^2$	change in degrees of freedom
1	small	2.40	-.43				-2
2	sum			.1			-1
3	"W"				1		0
4	$(sum)^2$					.01	-1
5	sum			-.025			0
6	small	.10	-.04				0

B) The STAGES

	total coefficients used: 1	small	sum	"W"	$(sum)^2$	spreads of residuals (in 0.01): midspread	8thspread	(*)	(**)
Before 1	0	0	0	0	0	89	140	120	296
After 1	2.40	-.43				54	79	46	97
After 2	2.40	-.43	.1			26	56	11	50
After 3	2.40	-.43	.1	1		26	51	11	41
After 4	2.40	-.43	.1	1	.01	16	45	4.1	30
After 5	2.40	-.43	.075	1	.01	15	35	3.6	20
After 6	2.50	-.47	.075	1	.01	13	41	2.7	27
(***)						(12)	(28)	(3.9)	(14)

(*) = $(\text{midspread})^2/(\text{degrees of freedom})$

(**) = $(\text{8thspread})^2/(\text{degrees of freedom})$

(***) After step 6 with 5 values set aside (5 more subtracted from degrees of freedom).

- a tilt on log smallest corner in the direction for biGaussian patterns, though apparently somewhat smaller -- size .43 or .47 as compared to .57. (The W pattern of biGaussian residuals may well be present).

- tilts on "sum" and "$(\text{sum})^2$" that combine to give the dependence shown in exhibit 41 -- relatively constant for large beans but substantially larger for small ones. Large and medium beans show similar strength of relationship; small beans show a greater strength.

29xG. Further analysis of other examples

The men's vision example, with only 3-by-3 = 9 double flogs and an unclear pattern of residuals (see exhibit 10) does not seem to require further analysis.

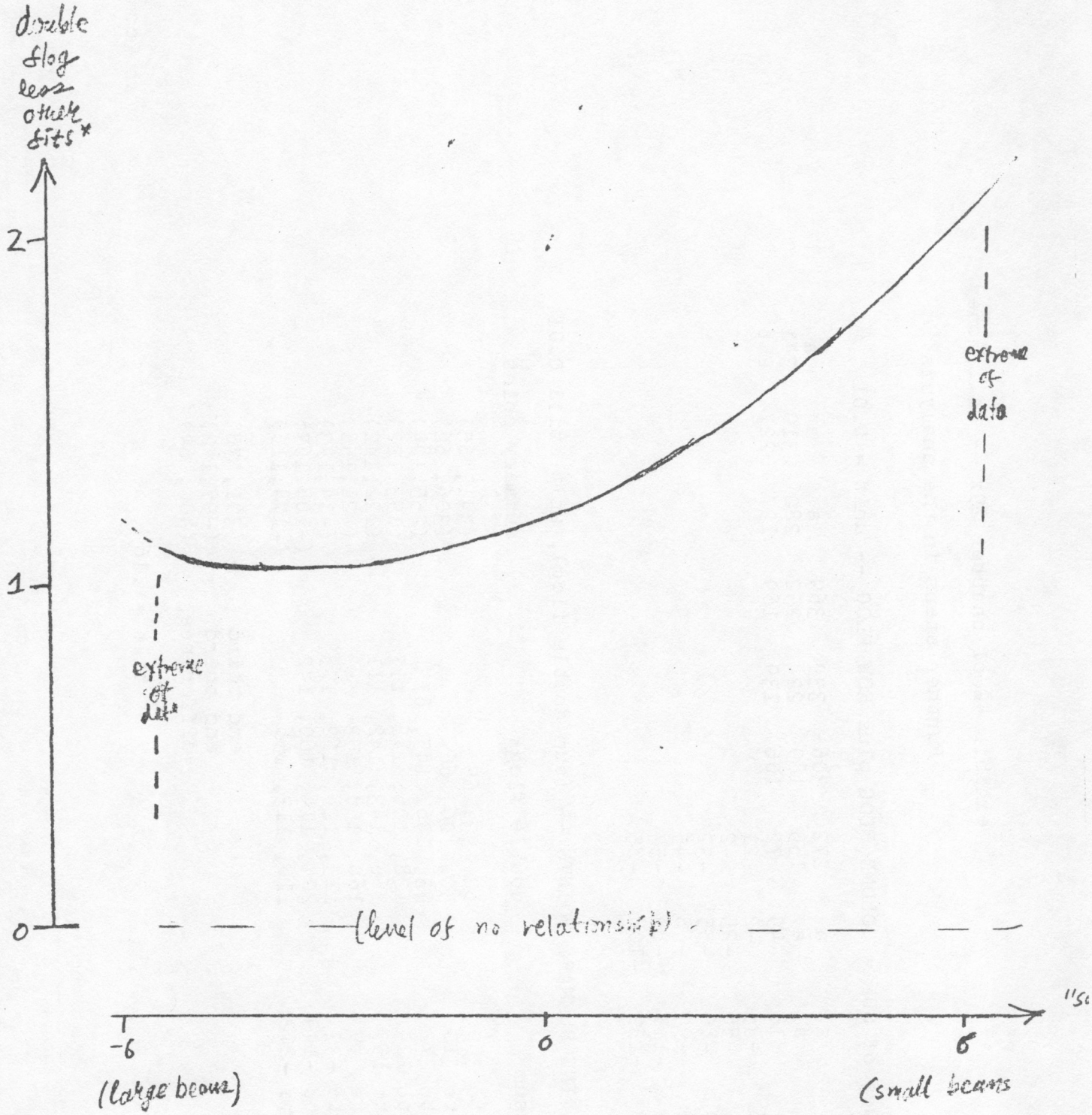
* call money *

The call-money example, which showed no tilt against log smallest corner, shows a SW-to-NE pattern (in panel D of exhibit 12). It is thus natural to look at its dependence on

sum = column flog PLUS raw flog

where we have chosen the signs of the flogs to increase from SW to NE. Exhibit 42 shows two steps of fitting, first in terms of sum and then in terms of $(\text{sum} - 1.22)^2$. (The choice of -1.22 makes vanish the entry for the middle summary point -- middle according to SUM.) As a result the midspread is

(The final fit against "sum"
("sum" = column flog + row flog)



* fit on log smallest column and W-shaped fit not included here

exhibit 42 of chapter 29x -- call money

Further steps in the analysis

A) Values of SUM = COLUMN FLOG plus ROW FLOG -- unit = 0.01

a	a	a	272	306	339	369	a	a	a
a	a	a	155	189	224	252	280	310	341
a	18	49	72	106	139	169	197	227	258
-96	-33	-2	21						
-119	-56	-25	-2						
-149	-86	-55	-32						
-162	-99	-68	-45						
-174	-111	-80	-57						
-202	-139	-108	-85						

B) The NINE SUMMARY POINTS -- (sum double flog), unit still 0.01

sum	double flogs	summary point
339 to 369	62, 93, 65	(341, 65)
280 to 310	30, 106, 66	(306, 66)
227 to 272	93, 146, 145, 67	(255,104)
109 to 224	128, 106, 148, 117	(193,122)
72 to 155	166, 143, 142, 111	(122,142)
18 to 49	160, 158, 153	(21,158)
-2 to -33	115, 124, 162, 175	(-14,143)
-56 to -149	204, 174, 160, 142	(-108,167)
-162 to -202	116, 113, 156	(-174,116)
	end third	(301, 78)
	end third	(-99,142)
	difference	(400, 64)

tilt = .16

(cont'd.)

C) Values of .16 times SUM

.	.	.	43	49	54	59	.	.	.
.	.		25	30	36	40	45	50	55
.	3	8	12	17	22	27	32	36	41
-15	-5	0	3						
-19	-9	-4	0						
-24									
-26									
-28									
-32									

D) Values of RESIDUALS = double flogs PLUS above MINUS 1.50

.	.	.	-14	-71	-34	2	.	.	.
.	.	.	41	8	-8	36	1	-34	-30
.	13	16	5	9	-17	25	-1	1	-42
39	-40	-26	6						
5	-17	8	25						
-14									
-60									
-65									
-16									

Median: 0. Hinges: -28 and 8h. 8ths: -41 and 25. Outs: 1/32

(cont'd.)

29x-92

exhibit 42 (cont'd.)

E) SUMMARY POINTS -- (sum,residuals) and ((sum-1.22)^2,residual), unit still 0.01

sum	residuals	summary point	summary point*
339 to 369	-34, 2, -30,	(341,-30)	(48,-30)
280 to 310	-71, 1, -34,	(306,-34)	(34,-34)
227 to 272	-14, 36, 1, -42	(255, -7)	(18, -7)
169 to 224	8, -8, 25, -1	(193, 4)	(5, 4)
72 to 155	41, 5, 9, -17	(122, 7)	(0, 7)
18 to 49	13, 16, 6	(21, 13)	(10, 13)
-2 to -33	-40, -26, 8, 25	(-14, -9)	(19, -9)
-56 to -149	39, 5, -14, -17	(-108, -4)	(53, -4)
-162 to -202	-60, -65, -17	(-174,-60)	(87,-60)
		top third	(63,-32)
		bottom third	(5, 10)
		difference	(58,-42)
		tilt = -.07	

* the (sum-1.22)^2 entries have unit = 0.1 not 0.01, hence also the tilt of -.07 instead of an apparent -.7.

(cont'd.)

29x-93

exhibit 42 (cont'd.)

F) Values of $-.07(\text{SUM}-1.22)^2$

.	.	.	-16	-24	0	-1			
.	.	.	-12	-3	-7	-12	-17	-25	-33
.	-7	-4	-2	0	0	-7	-4	8	-13
-33	-17	-11	0						
-41	-22	-15	-2						
-51									
-56									
-61									
-73									

G) Values of 2ND RESIDUAL = RESIDUAL minus ABOVE minus 8 -- unit still 0.01

.	.	.	-6	-48	-42	-5			
.	.	.	45	3	-9	40	10	-17	-5
.	12	6	-1	1	-25	19	-5	1	21
64	-23	-29	-2						
38	15	9	19						
29									
-12									
-12									
48									

Median: 1 Hinges: -10 and 19. 8ths: -27 and 39. Outs: 3/32

H) PROBLEM

Examine the relation of these 2nd residuals to log smallest corner (see exhibit 12).

29x-94

reduced from .53 to .36h to .29, while the 8thspread is reduced from .98 to .66 to .66. The first step is important; the second plausible.

* school-girls *

When we turn back to exhibit 7, we find rather small residuals. We have already exploited the small column effects by adding them back (in exhibit 8). What remains is to replace the row effects by a tilt on row flogs. Exhibit 43 shows the results:

The analysis leaves quite small residuals.

Though they are small, these residuals show the weak alternating behavior of columns quite clearly.

exhibit 43 of chapter 29x -- school girls

Approximation of earlier fit by a fit with only a few constants

A) FITTING a TILT on ROW FLOG -- unit 0.01

	flog	ff-log eff	.17 flog	diff	plus 6
α	-78	-18	-12	-5	1
β	-1	-8	0	-8	-2
γ	59	6	10	-4	2
δ	110	12	19	-7	-1

"fflog eff" column from panel A of exhibit 8 by subtraction of a uniform 44.

B) RESULT of FURTHER ADDING BACK -- unit still 0.01

	Residuals			
	λ	μ	ν	π
α	-3	4	-3	13
β	-6	2	-2	-3
γ	13	6	-4	2
δ	1	2	-9	0

Median: 1h. Hinges: -3h and 3h. 8ths: -5 and 9h. Outs: 2/16.

Fit = .21(row flog-.29)(column flog-.29) + .17(row flog) + .38.

****

29xH. <u>How far have we come?</u>

In this chapter we have gone ahead substantially. One area of progress can be thus summarized:

- we have found out how to take apart a table of counts according to two ordered responses (doubly ordered true contingency table). (This is a "taking apart" in the sense that from the pieces we can recover the original counts.)

- in doing this we have emphasized cuts -- single and double -- rather than cells or stripes.

- our breakdown gives a total, one single flog for each single cut and one double flog for each double cut.

- if the table of counts has no zeros, our breakdown has in all exactly as many numbers as there were counts.

- we have found, in a variety of examples, that the table of double flogs -- which is an ordinary 2-way table without any count-like properties -- can be further analyzed in various interesting ways.

In doing this, our key tool was the double flog, a double folding (double differencing) of logs, which takes the form

log (NE count) - log (NW count) - log (SE count) + log (SW count)

for a two-by-two table divided N-S and E-W. We have used both "raw double flogs", as above, and "matched double flogs", which are 1.15 times as large (and related conveniently to the matched (single) flogs of chapter 28).

With this under our belt, we turned our attention to the

log smallest corner

which is the smallest -- whichever corner that may involve -- of the four log counts that enter a double flog. This quantity entered first as a rough indication of the relative sensitivity of different double flogs to small changes in the counts defining it. Almost immediately it gained another kind of usefulness. In summary:

- many 2-way tables of double flogs show a tilt carried by log smallest corner.
- sometimes this tilt is positive, sometimes negative. (Sometimes it seems to be absent.)
- if we take any one of the only widely mentioned theoretical 2-way patterns of distribution -- the bivariate Gaussian distribution here called biGaussians -- we find that the corresponding double flogs, when plotted against log smallest corner, are surprisingly close to a straight line of negative slope. (From the residuals from this

straight line we can extract a further, smaller dependence on "column flog - row flog".)

- accordingly we can examine 2-way tables of double flogs as to one aspect of their possible relation to biGaussian patterns of distribution by relating their tilt against log smallest corner and their level to what happens for actual biGaussians. Some examples seem to approach matching, others are very far away.

Having begun to fit 2-way tables of double flogs with a tilt against a carrier, it is natural to go further. Quite limited exploration suggests that we give particular attention to such carriers as:

log smallest corner

column flog - row flog

column flog + row flog

the squares of the last two carriers (see also below).

Sometimes it seems appropriate to make a column-plus-row analysis of a 2-way table of double flogs. When we do this, we will do well to bear in mind that:

- restraining the column-plus-row fit, perhaps to

(coeff)(column flog) + (coeff')(row flog)

is often worthwhile.

- the natural extension of a column-plus-row fit uses the product carrier based on single flogs

(column flog) times (row flog)

and NOT a product carrier based on stripe effects for the double flogs.

- this product carrier can also be written as

$$\frac{1}{4}\{(\text{column flog}) \text{ plus } (\text{row flog})\}^2 - \frac{1}{4}\{(\text{column flog}) \text{ minus } (\text{row flog})\}^2$$

so that appearance of the squares of both sum and difference with nearly **opposite** coefficients always raises the question of whether we should look at the product carrier.

- for patterns of distribution resembling the biGaussian at least, we can often fit log smallest corner quite well with a straight line against

$$\{(\text{column flog}) \text{ minus } (\text{row flog})\}^2 + (\text{coeff})\{(\text{column flog}) \text{ plus } (\text{row flog})\}^2$$

where the coefficient is to be learned about in problems 11, 13, 14 and 16. Thus the appearance of the squares of the sum and difference with nearly the **same** coefficients always raises the question of whether we should look at log smallest corner. Exchanges in the other direction have also to be kept in mind.

- combinations of the last can represent diverse quadratics.

All in all, we have gained powerful tools for dealing with ordered contingency tables -- or with the two-way portion of bivariate distributions.

Based on quite limited experience, we have also gained strong doubt of the general applicability of:

- biGaussian patterns of distribution
- any simple family of standard patterns of bivariate distribution, where any family with only 1 relationship parameter is surely "simple", as are many with 2 or 3.

****

29xP. Problems for chapter 29x

1. Do the problem of exhibit 2.
2. Do problem 1 of exhibit 4.
3. Do problem 2 of exhibit 4.
4. Do the problem of exhibit 5.
5. Do the problem of exhibit 8.
6. Do the problem of exhibit 9.
7. Do problem 1 of exhibit 12.
8. Do problem 2 of exhibit 12.
9. Do the problem of exhibit 14.
10. Do problem 1 of exhibit 21.
11. Do problem 2 of exhibit 21.
12. Do problem 1 of exhibit 22.
13. Do problem 2 of exhibit 22.
14. Do problem 1 of exhibit 23.
15. Do problem 2 of exhibit 23.
16. Do the problem of exhibit 24.
17. Do the problem of exhibit 27.
18. Do the problem of exhibit 28.
19. Do the problem of exhibit 30.
20. Do the problem of exhibit 35.
21. Do the problem of exhibit 39.
22. Do the problem of exhibit 42.

23. Kendall and Stuart (1961. The Advanced Theory of Statistics. Table 335 at page 586 of volume 2 of 3.) also give counts according to unaided distance vision of women employed by the U.K. Royal Ordnance factories in 1943-46. For left eye strengths (highest, second, third, lowest; (total)) the counts are: right eye highest(1520,266,124,66; (1976)), right eye second (234,1512,432,78;(2256)), right eye third(117,362,1772,205;(2456)), right eye lowest (36,82,179,492,(789)), and totals (1907,2220,2507,841,(7477)). Carry out an analysis in the style of section 29xC.

29xR. Technical references

Frederick Mosteller 1968. Association and estimation in contingency tables. J. American Statistical Association 63: 1-28.

Has tables of $e^{(ff-log)}$ for quadrants on page 5.

Table 4 on p. 6, varying ρ to give matched double-flog = .345.

R. L. Plackett 1965. A class of bivariate distributions. J. American Statistical Association 60: 516-522.

Case of constant double-flog. Compares matched double-flog = .69 with ρ = .5 for Gaussian margins (Table 1 on page 519) for 0.5 × 0.5 cells. Also conditional distribution.